Third edition

Chomsky's Universal Grammar

AN INTRODUCTION

Vivian Cook and Mark Newson

Blackwell Publishing

Chomsky's Universal Grammar

An Introduction

Third Edition

V. J. Cook and Mark Newson

Blackwell
Publishing

© 2007 by V. J. Cook

BLACKWELL PUBLISHING
350 Main Street, Malden, MA 02148-5020, USA
9600 Garsington Road, Oxford OX4 2DQ, UK
550 Swanston Street, Carlton, Victoria 3053, Australia

The right of V. J. Cook and Mark Newson to be identified as the Authors of this Work has been asserted in accordance with the UK Copyright, Designs, and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs, and Patents Act 1988, without the prior permission of the publisher.

First edition published 1988 by Basil Blackwell Ltd
Second edition published 1996 by Blackwell Publishers Ltd
Third edition published 2007 by Blackwell Publishing Ltd

1 2007

Library of Congress Cataloging-in-Publication Data

Cook, V. J. (Vivian James), 1940–
Chomsky's universal grammar: an introduction / V. J. Cook and Mark Newson. — 3rd ed.
p. cm.
Includes bibliographical references and index.
ISBN-13: 978-1-4051-1186-7 (hardcover : alk. paper)
ISBN-13: 978-1-4051-1187-4 (pbk. : alk. paper) 1. Grammar, Comparative and general. 2. Language acquisition. 3. Chomsky, Noam. I. Newson, Mark. II. Title.
P151.C64 2007
415—dc22

2006036117

A catalogue record for this title is available from the British Library.

Set in 10/12.5pt Palatino
by Graphicraft Limited, Hong Kong
Printed and bound in Singapore
by C.O.S. Printers Pte Ltd

The publisher's policy is to use permanent paper from mills that operate a sustainable forestry policy, and which has been manufactured from pulp processed using acid-free and elementary chlorine-free practices. Furthermore, the publisher ensures that the text paper and cover board used have met acceptable environmental accreditation standards.

For further information on
Blackwell Publishing, visit our website:
www.blackwellpublishing.com

Contents

Preface to the Third Edition vii

1 The Nature of Universal Grammar 1
1.1 The early development of Universal Grammar Theory 2
1.2 Relating 'sounds' and 'meanings' 4
1.3 The computational system 8
1.4 Questions for linguistics 11
1.5 General ideas of language 13
1.6 Linguistic universals 20
1.7 The evidence for Universal Grammar Theory 24
1.8 Conclusion 26

2 Principles, Parameters and Language Acquisition 28
2.1 Principles and parameters 28
2.2 Language acquisition 45

3 Structure in the Government/Binding Model 61
3.1 The heart of the Government/Binding Model 62
3.2 Modules, principles and parameters 62
3.3 X-bar Theory in Government and Binding 73
3.4 Theta Theory 80
3.5 Control Theory and null subjects 86
3.6 Further developments in X-bar Theory 100
3.7 Summary 118

4 Movement in Government/Binding Theory 121
4.1 An overview of movement 121
4.2 Further developments to the theory of movement 133
4.3 Bounding, Barriers and Relativized Minimality 139
4.4 Case Theory 146
4.5 Binding Theory 162
4.6 Beyond S-structure and the Empty Category Principle 175

5 Chomskyan Approaches to Language Acquisition **185**
5.1 The physical basis for Universal Grammar 185
5.2 A language learning model 189
5.3 The innateness hypothesis 204
5.4 The role of Universal Grammar in learning 205
5.5 Complete from the beginning or developing with time? 207
5.6 Issues in parameter setting 209
5.7 Markedness and language development 215

6 Second Language Acquisition and Universal Grammar **221**
6.1 The purity of the monolingual argument 221
6.2 Universal bilingualism 222
6.3 The multi-competence view 223
6.4 The poverty-of-the-stimulus argument and second language acquisition 224
6.5 Models and metaphors 228
6.6 Hypotheses of the initial second language state 231
6.7 The final state of second language acquisition 238

7 Structure in the Minimalist Program **242**
7.1 From Government/Binding to the Minimalist Program 243
7.2 Basic minimalist concepts 249
7.3 Phrase structure in the Minimalist Program 255
7.4 Thematic roles and structural positions 262
7.5 Adjunction 265
7.6 Linear order 268

8 Movement in the Minimalist Program **271**
8.1 Functional heads and projections 271
8.2 The motivation for movement 275
8.3 The nature of movement 279
8.4 Overt and covert movement 281
8.5 Properties of movement 287
8.6 Phases 301
8.7 Conclusion 308

References 310
Index 319

1 The Nature of Universal Grammar

The idea of Universal Grammar (UG) put forward by Noam Chomsky has been a crucial driving force in linguistics. Whether linguists agree with it or not, they have defined themselves by their reactions to it, not only in terms of general concepts of language and language acquisition, but also in how they carry out linguistic description. From the 1960s to the 1980s, UG became a flash-point for disciplines outside linguistics such as psychology, computer parsing of language and first language acquisition, even if these areas have tended to lose contact in recent years. The aim of this book is to convey why Chomsky's theories of language still continue to be stimulating and adventurous and why they have important consequences for all those working with language.

This book is intended as an introduction to Chomsky's UG Theory for those who want a broad overview with sufficient detail to see how its main concepts work rather than for those who are specialist students of syntax, for whom technical introductions such as Adger (2003) and Hornstein et al. (2005) are more appropriate. Nor does it cover Chomsky's political views, still as much a thorn in the side of the US establishment as ever, for example Chomsky (2004a). While the book pays attention to the current theory, called the Minimalist Program, it concentrates on providing a background to the overall concepts of Chomsky's theory, which have unfolded over six decades. Where possible, concepts are illustrated through Chomsky's own words. The distinctive feature of the book is the combination of Chomsky's general ideas of language and language acquisition with the details of syntax.

This opening chapter sets the scene by discussing some of the general issues of Chomsky's work on the notion of UG. Following this, chapter 2 discusses central concepts of the framework and how these relate to Chomsky's views on language acquisition. The next two chapters provide an introduction to the syntax of Government/Binding Theory in terms of structure and of movement respectively. Chapter 5 looks at Chomskyan approaches to first language acquisition, chapter 6 at second language acquisition. Then chapters 7 and 8 outline the current Minimalist Program, again separating structure and movement.

Two conventions followed in this book need briefly stating. As usual in linguistics books, an asterisk indicates an ungrammatical sentence. Example sentences, phrases and structures are numbered for ease of reference, i.e.:

(1) *That John left early seemed.

While much of the discussion is based upon English for convenience, the UG Theory gains its power by being applied to many languages. Indeed the past twenty years have seen a proliferation in the languages studied, which will be drawn on when possible. It should perhaps be pointed out that the sentences used in this book are examples of particular syntactic issues rather than necessarily being based on complete recent analyses of the languages in question.

1.1 The early development of Universal Grammar Theory

The approach adopted in this book is to look at the general ideas of the Chomskyan theory of UG without reference to their historical origins. Nevertheless some allusions have to be made to the different versions that have been employed over the years and the history of the theory needs to be briefly sketched, partly so that the reader is not confused by picking up a book with other terminology.

Development has taken place at two levels. On one level are the general concepts about language and language acquisition on which the theory is based. The origins of such ideas as competence and performance or the innateness of language can be traced back to the late fifties or mid-sixties. These have grown continuously over the years rather than being superseded or abandoned. On this level the UG Theory is recognizable in any of its incarnations and the broad outlines have remained substantially the same despite numerous additions.

On another level come ideas about the description of syntax, which fall into definite historical phases. Different periods in the Chomskyan description of syntax have tended to become known by the names of particular books. Each was characterized by certain concepts, which were often rejected by the next period; hence the statements of one period are often difficult to translate into those of the next. Unlike the continuity of the general ideas, there are shifts in the concepts of syntax, leading to a series of apparent discontinuities and changes of direction.

The original model, **Syntactic Structures**, took its name from the title of Chomsky's 1957 book, which established the notion of 'generative grammar' itself, with its emphasis on explicit 'generative', formal description through 'rewrite rules' such as S → NP VP, as described below. It made a separation between phrase structure rules that generated the basic structures, called 'kernel sentences', and transformations which altered these in various ways by turning them into passive or negative sentences etc.; hence its popular name was 'transformational generative grammar' or 'TGG'. Its most memorable product was the sentence:

(2) Colourless green ideas sleep furiously.

intended to demonstrate that sentences could be grammatical but meaningless and hence that syntax is independent of semantics. This sentence became so widely known that attempts were made to create poems that included it naturally (after all, Andrew Marvell wrote of 'a green thought in a green shade').

This theory was superseded by the model first known as the **Aspects Model** after Chomsky's 1965 book *Aspects of the Theory of Syntax*, later as the **Standard Theory**. This was distinguished by the introduction of the competence/performance distinction between language knowledge and language use and by its recognition of 'deep' and 'surface' structure in the sentence. Two classic test sentences were:

(3) John is eager to please.

which implies that John pleases *other people*, and:

(4) John is easy to please.

which implies that other people please *John*. This difference is captured by claiming that the two sentences have the same surface structure but differ in deep structure, where *John* may act as the subject or object of the verb *please*. Again, sentence (4) was so widely known it featured in graffiti on London street walls in the 1970s and was used as a book title (Mehta, 1971).

During the 1970s the Standard Theory evolved into the **Extended Standard Theory (EST)**, which refined the types of rules that were employed. This in turn changed more radically into the **Government/Binding (GB) Model** (after *Lectures on Government and Binding*, Chomsky, 1981a), which substantially underpins this book. The GB Model claimed that human languages consisted of principles that were the same for any grammar and parameters that allowed grammars to vary in limited ways, to be illustrated in the next chapter. It also refined deep and surface structure into the more technical notions of 'D-structure' and 'S-structure', to be discussed below. The GB version of UG was presented most readably in *Knowledge of Language* (Chomsky, 1986a). Though 'Government and Binding Theory' was the common label for this model, Chomsky himself found it misleading because it gave undue prominence to two of its many elements: 'these modules of language stand alongside many others . . . Determination of the nature of these and other systems is a common project, not specific to this particular conception of the nature of language and its use' (Chomsky, 1995a, pp. 29–30). Hence the label of **Principles and Parameters (P&P) Theory** has come to be seen as closer to its essence, and can still be applied to the contemporary model.

Since the late eighties a further major model of syntax has been undergoing development, a model called the **Minimalist Program (MP)**, again reflected in the title of Chomsky's first publication in this framework (Chomsky, 1993) and his later book (Chomsky, 1995a). So far this has had three phases. In the first phase, up till about 1996, the MP concentrated on the general features of the model, simplifying knowledge of language to invariant principles common to all languages, and, by attaching parameters to the vocabulary, making everything that people have to acquire in order to know a particular language part of the lexicon. From about 1996 the second phase embarked on a programme of radically rethinking syntax, eliminating much of the apparatus of GB Theory in favour of a minimal set of operations and ideas and exploring whether the central 'computational system' of language interfaces 'perfectly' with phonology and cognition. Since 2000

Starting date	Model	Key terms	Key book/article
1957	Transformational generative grammar (TG)	Rewrite rules Transformation Generative Kernel sentence	Chomsky, 1957
1965	Aspects, later Standard Theory	Competence/performance Deep/surface structure	Chomsky, 1965
c. 1970	Extended Standard Theory (EST)		Chomsky, 1970
1981	Government/Binding Theory (GB)	Principles Parameters D- and S-structure Movement	Chomsky, 1981a
post-1990	Minimalist Program (MP)	Computational system Interface conditions Perfection	Chomsky, 1993

Figure 1.1 Phases in the development of Chomsky's Universal Grammar

a new model has been emerging, known as the **Phases Model**. A current view is presented in chapters 7 and 8 below. A readable set of lectures by Chomsky on this framework is *The Architecture of Language* (Chomsky, 2000a).

1.2 Relating 'sounds' and 'meanings'

The focus of all Chomsky's theories has been what we know about language and where this knowledge comes from. The major concern is always the human mind. The claims of UG Theory have repercussions for how we conceive of human beings and for what makes a human being. Language is seen as something in the individual mind of every human being. Hence it deals with general properties of language found everywhere rather than the idiosyncrasies of a particular language such as English or Korean – what is common to human beings, not what distinguishes one person from another. Everybody knows language: what is it that we know and how did we come to acquire it?

As well as this invisible existence within our minds, language also consists of physical events and objects, whether the sounds of speech or the symbols of writing: language relates to things outside our minds. The fundamental question

for linguistics since Aristotle has been how language bridges the gap between the intangible interior world of knowledge and the concrete physical world of sounds and objects; 'each language can be regarded as a particular relationship between sounds and meaning' (Chomsky, 1972a, p. 17). So a sentence such as:

(5) The moon shone through the trees.

consists on the one hand of a sequence of sounds or letters, on the other of a set of meanings about an object called 'the moon' and a relationship with some other objects called 'trees'. Similarly the Japanese sentence:

(6) Ohayoh gozaimasu.

is connected to its Japanese pronunciation on the one side and to its meaning 'Good morning' on the other.

The sounds and written symbols are the external face of language, its contact with the world through physical forms; they have no meaning in themselves. *Moon* means nothing to a monolingual speaker of Japanese, *gozaimasu* nothing to a monolingual English speaker. The meanings are the internal face of language, its contact with the rest of cognition; they are abstract mental representations, independent of physical forms. The task of linguistics is to establish the nature of this relationship between external sounds and internal meanings, as seen in figure 1.2.

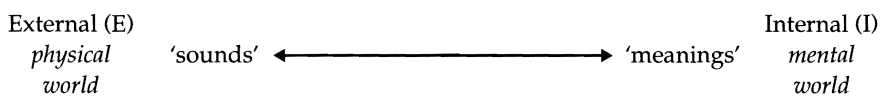


Figure 1.2 The sound–meaning link

If language could be dealt with either as pure sounds or as pure meanings, its description would be simple: *moon* is pronounced with phonemes taken from a limited inventory of sounds; *moon* has meanings based on a limited number of concepts. The difficulty of handling language is due to the complex and often baffling links between them: how *do* you match sounds with meanings? How does *The moon shone through the trees* convey something to an English speaker about a particular happening?

According to Chomsky (for instance Chomsky, 1993), the human mind bridges this gap via a 'computational system' that relates meanings to sequences of sounds in one direction and sequences of sounds to meanings in the other.

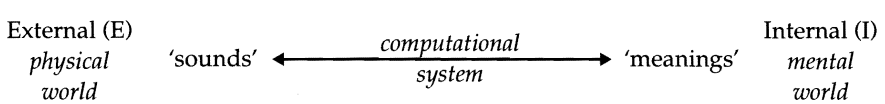


Figure 1.3 The computational system

The sheer sounds of language, whether produced by speakers or perceived by listeners, are linked to the meanings in their minds by the computational system.

What speakers of a language know is not just the sounds or the meanings but how to connect the two. The complexity of language resides in the features of the computational system, primarily in the syntax.

Since the late 1990s, as part of the MP, Chomsky has been interested in exploring the connections between the central computational system and, on the one side, the physical expressions of language, on the other, the mental representation of concepts: what happens at the points of contact between the computational system and the rest? At the point of contact with sounds, the mind needs to change the internal forms of language used by the computational system into actual physical sounds or letters through complex commands to muscles, called by Chomsky (2001a) the 'sensorimotor system'; i.e. *the moon* is said using the appropriate articulations. In reverse, the listener's mind has to convert sounds into the forms of representation used by the computational system, so that is perceived as the phrase *the moon*.

At the point of contact with meanings, the mind needs to change the representation of language used by the computational system into the general concepts used by the mind, called 'the conceptual-intentional system' (Chomsky, 2001a), i.e. *moon* is connected to the concept of 'earth's satellite'. Going in the opposite direction, while speaking the mind has to convert the concepts into linguistic representation for the computational system, i.e. 'earth's satellite' is converted into *moon*. Figure 1.4 incorporates these interfaces into the bridge between sounds and meanings.

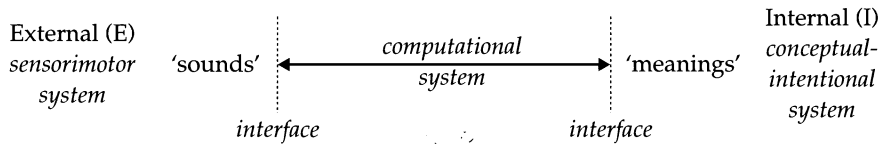


Figure 1.4 The interfaces with the computational system

The points of contact are the interfaces between the computational system, that is to say, between language knowledge, and two other things which are *not* language – the outside world of sounds and the inside world of concepts. For the computational system to work, it must be able to interface in both directions, to have two 'access systems' (Chomsky, 2000a, p. 9). Or, put another way, language is shaped by having to be expressed on the one hand as sounds or letters that can be handled by the human body, on the other as concepts that can be conceived by the human mind.

Let us now add some of the apparatus of principles and parameters to this picture. To describe a sentence such as:

(7) Gill teaches physics.

the grammar must show how the sentence is pronounced – the sequence of sounds, the stress patterns, the intonation, and so on; what the sentence actually means – the individual words – *Gill* is a proper name, usually female, and so on; and how these relate to one another via syntactic devices such as the subject (*Gill*)

coming before the verb (*teaches*), with the object (*physics*) after. The term 'grammar' is used generally to refer to the whole knowledge of language in the person's mind rather than just to syntax, which occasionally leads to misinterpretation by outsiders. The linguist's grammar thus needs a way of describing actual sounds – a phonetic representation; it needs a way of representing meaning – a semantic representation; and it needs a way of describing the syntactic structure that connects them – a syntactic level of representation. Syntactic structure plays a central mediating role between physical form and abstract meaning.

Principles and Parameters Theory captures this bridge between sound and meaning through the technical constructs **Phonetic Form (PF)**, realized as sound sequences, and **Logical Form (LF)**, representations of certain aspects of meaning, connected via the computational system, as shown in figure 1.5:

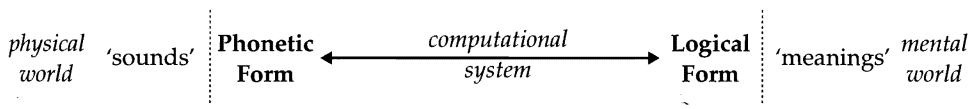


Figure 1.5 The computational system

PF and LF have their own natures, for which distinct PF and LF components are needed within the model. They form the contact between the grammar and other areas, at the one end physical realizations of sounds, at the other further mental systems: 'PF and LF constitute the "interface" between language and other cognitive systems, yielding direct representation of sound on the one hand and meanings on the other as language and other systems interact' (Chomsky, 1986a, p. 68).

Most principles-and-parameters-based research has concentrated on the central computational component rather than on PF or LF. If syntax is a bridge, independent theories of PF or LF are beside the point: however elegant the theories of PF or LF may be in themselves, they must be capable of taking their place in the bridge linking sounds and meanings. The same is true of language acquisition: the central linguistic problem is how the child acquires the elements of the computational system rather than the sounds or meanings of the language. PF and LF are treated in this book as incidentals to the main theme of syntax. Throughout the development of Chomsky's models, a key aspect has been their 'syntactocentrism' (Jackendoff, 2002): syntax has always been the key element of knowledge of language. Perhaps this is why the word 'grammar' is often extended in Chomskyan theories to encompass the whole knowledge of language in the individual's mind.

Nevertheless this does not mean that considerable work on theories of LF and PF has not been carried out over the years. The PF component for example grew from *The Sound Pattern of English* (Chomsky and Halle, 1968) into a whole movement of generative phonology, as described in Roca (1994) and Kenstowicz (1994).

The bridge between sounds and meanings shown in figure 1.5 is still not complete in that LF represents essentially 'syntactic' meaning. 'By the phrase "logical form" I mean that partial representation of meaning that is determined by grammatical structure' (Chomsky, 1979, p. 165). LF is not in itself a full semantic representation but represents the structurally determined aspects of meaning that form

one input to a semantic representation, for example the difference in interpreting the direction:

(8) It's right opposite the church.

as:

(9) It's [right opposite] the church.

meaning 'exactly opposite the church', or as:

(10) It's right [opposite the church].

meaning 'turn right when you are opposite the church'.

The simplest version of linguistic theory needs two levels to connect the computational system with the physical articulation and perception of language on the one hand and with the cognitive semantic system on the other. The MP indeed claims that 'a particularly simple design for language would take the (conceptually necessary) interface levels to be the only levels' (Chomsky, 1993, p. 3).

1.3 The computational system

The make-up of the computational system is then the subject-matter of linguistics. One vital component is the lexicon in the speaker's mind containing all their vocabulary, analogous to a dictionary. This knowledge is organized in 'lexical entries' for each word they know. While the meaning of the sentence depends upon the relationships between its various elements, say how *the moon* relates to *shone*, it also depends upon the individual words such as *moon*. We need to know the diverse attributes of the word – that it means 'satellite of a planet', that it is pronounced [mu:n], that it is a countable noun, *a moon*, and so on. Each lexical entry in the mental lexicon contains a mass of information about how the word behaves in sentences as well as its 'meaning'. Our knowledge of language consists of thousands of these entries encoding words and their meanings. The computational system relies upon this mental lexicon. Since the early 1980s Chomskyan theories have tied the structure of the sentence in with the properties of its lexical items rather than keeping syntax and the lexicon in watertight compartments. To a large extent the choice of lexical items drives the syntax of the computational system, laying down what structures are and aren't possible in a sentence having those lexical items: if you choose the verb *read*, you've got to include an object (*read something*); if you choose the noun *newspaper* in the singular, you've got to have a determiner (*a/the newspaper*).

The second vital component in the computational system is the principles of UG. Knowledge of language is based upon a core set of principles embodied in all languages and in the minds of all human beings. It doesn't matter whether one speaks English, Japanese or Inuktitut: at some level of abstraction all languages

rely on the same set of principles. The differences between languages amount to a limited choice between a certain number of variables, called parameters. Since the early 1980s the Chomskyan approach has concentrated on establishing more and more powerful principles for language knowledge, leading to the MP. The central claim that language knowledge consists of principles universal to all languages and parameters whose values vary from one language to another was the most radical breach with the view of syntax prevailing before 1980, which saw language as 'rules' or 'structures' and believed language variation was effectively limitless.

Principles apply across all areas of language rather than to a single construction and they are employed wherever they are needed. Knowledge of language does not consist of rules as such but of underlying principles from which individual rules are derived. The concept of the rule, once the dominant way of thinking about linguistic knowledge, has now been minimized. 'The basic assumption of the P&P model is that languages have no rules at all in anything like the traditional sense, and no grammatical constructions (relative clauses, passives, etc.) except as taxonomic artifacts' (Chomsky, 1995b, p. 388). The change from rules to principles was then a major development in Chomskyan thinking, the repercussions of which have not necessarily been appreciated by those working in psychology and other areas, who still often assume Chomskyan theory to be rule-based. 'There has been a gradual shift of focus from the study of rule systems, which have increasingly been regarded as impoverished, . . . to the study of systems of principles, which appear to occupy a much more central position in determining the character and variety of possible human languages' (Chomsky, 1982, pp. 7–8). The information stated in rules has to be reinterpreted as general principles that affect all rules rather than as a property of individual rules. Rules are by-products of the interaction between the principles and the lexicon. UG Theory is not concerned with specific syntactic constructions such as 'passive' or 'relative clause' or 'question', or the rules which linguists can formulate to express regularities in them, which are simply convenient labels for particular interactions of principles and parameters. The passive is not an independent construction so much as the product of a complex interaction of many principles and parameter settings, each of which also has effects elsewhere in the syntax: 'a language is not, then, a system of rules, but a set of specifications for parameters in an invariant system of principles of Universal Grammar (UG)' (Chomsky, 1995b, p. 388).

Figure 1.6 then incorporates the principles and the lexicon into the computational system. The lexicon is the key starting point for a sentence; the principles combine with the properties of the lexical items chosen to yield a representation

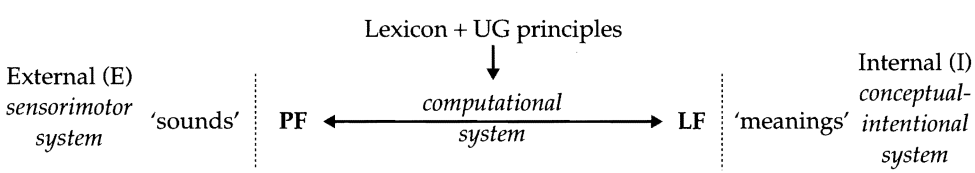


Figure 1.6 The computational system

that is capable of connecting with the sounds and meanings outside the computational system.

EXERCISE 1.1 Here are some 'rules' for everyday human behaviour in different parts of the world. Could these be seen as examples of more general principles? How are these different from rules and principles of language?

Wash your hands in between the courses of meals.
Drive on the left of the road.
Add salt only after potatoes have boiled.
Do not use a capital letter after a colon.
The evening meal should be eaten about 11 p.m.
Children should be seen but not heard.
Women must wear a head scarf.
For examinations students must wear a dark suit, a gown and a mortar board.

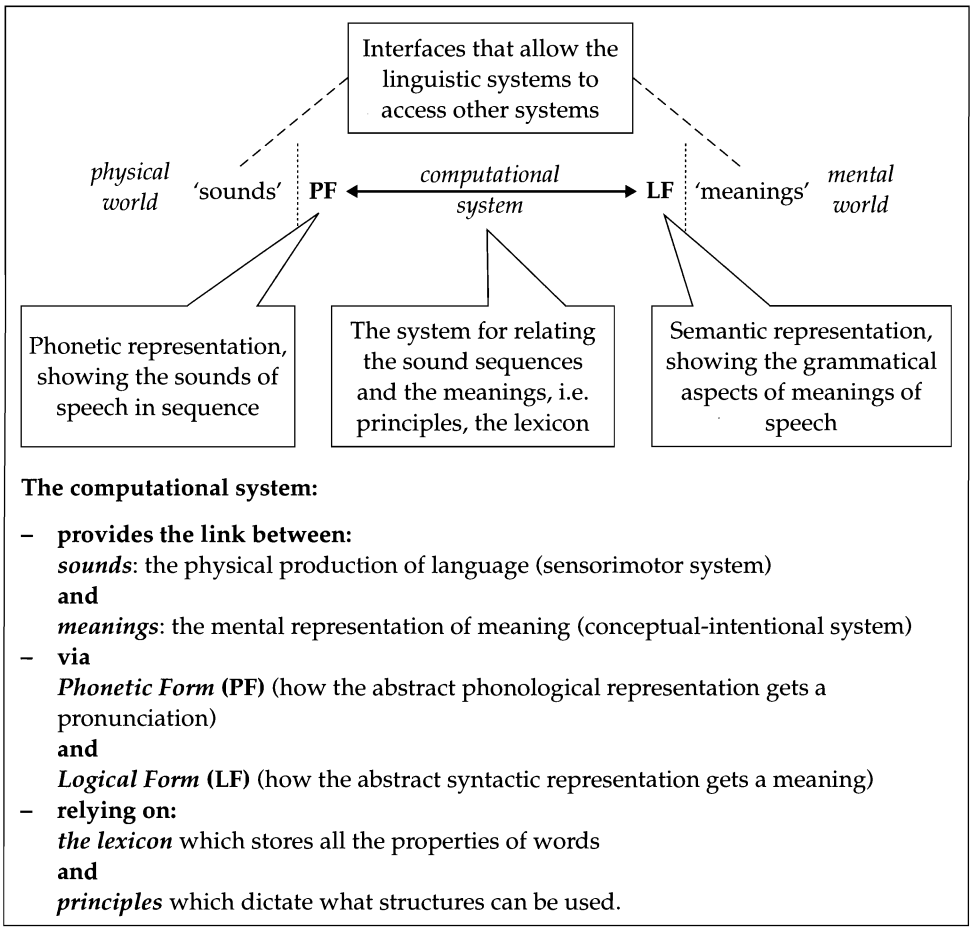


Figure 1.7 Explanatory diagram of the computational system

1.4 Questions for linguistics

To sum up what has been said so far, UG is a theory of knowledge that is concerned with the internal structure of the human mind – how the computational system links sounds to meaning. Since the early 1980s it has claimed that this knowledge consists of a set of principles that apply to all languages and parameters with settings that vary between languages. Acquiring language therefore means learning how these principles apply to a particular language and which value is appropriate for each parameter for that language. Each principle or parameter that is proposed is a substantive claim about the mind of the speaker and about the nature of language acquisition, not just about the description of a single language. UG Theory is making precise statements about properties of the mind based on specific evidence, not vague or unverifiable suggestions. The general concepts of the theory are inextricably tied to the specific details; the importance of UG Theory is its attempt to integrate grammar, mind and language at every moment.

The P&P approach is not, however, tied to a particular model of syntactic description; 'it is a kind of theoretical framework, it is a way of thinking about language' (Chomsky, 2000a, p. 15). Historically speaking it arose out of Chomsky's work of the early 1980s, notably Chomsky (1981a). The MP of the 1990s and 2000s has remained within an overall P&P 'framework' despite abandoning, modifying or simplifying many of them. P&P Theory is a way of thinking about knowledge of language as consisting of certain fixed and constant elements and some highly restricted variable elements, and so it can be implemented in different ways. The MP intends to cut down the number of operations and assumptions, making it in the end simpler than past theories.

The aims of linguistics are often summed up by Chomsky, for example Chomsky (1991a), in the form of three questions:

(i) *What constitutes knowledge of language?*
The linguist's prime duty is to describe what people know about language – whatever it is that they have in their minds when they know English or French or any language, or, more precisely, a grammar. Speakers of English know, among many other things, that:

(11) Is Sam the cat that is black?

is a sentence of English, while:

(12) *Is Sam is the cat that black?

is not, even if they have never met a sentence like (12) before in their lives.

(ii) *How is such knowledge acquired?*
A second aim for linguistics is to discover how people acquire this knowledge of language. Inquiring what this knowledge is like cannot be separated from asking

how it is acquired. Anything that is proposed about the nature of language knowledge begs for an explanation of how it came into being. What could have created the knowledge in our minds that sentence (11) is possible, sentence (12) is impossible? Since the knowledge is present in our minds, it must have come from somewhere; the linguist has to show that there is in principle some source for it. This argument is expanded on a grander scale in chapter 5. Logically speaking, explaining the acquisition of language knowledge means first establishing what the knowledge that is acquired actually consists of, i.e. on first answering question (i).

(iii) *How is such knowledge put to use?*
A third aim is to see how this acquired language knowledge is actually used. Sentence (11) presumably could be used to distinguish one cat out of many in a photograph. Again, investigating how knowledge is used depends on first establishing what knowledge *is*, i.e. on the answer to question (i).

Sometimes a fourth question is added, as in Chomsky (1988, p. 3):

(iv) *What are the physical mechanisms that serve as the material basis for this system of knowledge and for the use of this knowledge?*
This mental knowledge must have some physical correlate, in other words there must be some relationship between the computational system in the mind and the physical structures of the brain. The principles and parameters of UG themselves must be stored somewhere within the speaker's brain. Though our understanding of the physical basic for memory is advancing rapidly with the advent of modern methods of brain scanning, Chomsky (2005b, p. 143) points out 'current understanding falls well short of laying the basis for the unification of the sciences of the brain and higher mental faculties, language among them'. Useful accounts of such brain research can be found in Fabbro (1999) and Paradis (2004). It will not be tackled in this book.

Recently Chomsky has begun to pose more difficult questions about the nature of language and the extent to which this is determined by restrictions imposed on it from the other systems which interpret it: 'How good a solution is language to certain boundary conditions that are imposed by the architecture of the mind?' (Chomsky, 2000a, p. 17). Pre-theoretically, the linguistic system might be a completely independent aspect of the human mind which happens to be made use of by the sensorimotor system and the conceptual-intentional system, but which has distinct properties all of its own. On the other hand, the linguistic system might be an optimal solution to the problem of bridging the gap between the interpretative systems which has no properties other than those necessary for completion of its function, this being to deliver objects which are 'legible' at the interface levels, i.e. which meet the 'legibility conditions' imposed by the sensorimotor and conceptual-intentional systems. Chomsky discusses this in terms of a concept of perfection, the discovery of which he sees as being at the heart of science; 'language is surprisingly close to perfect in that very curious sense; that is, it is a near-optimal solution to the legibility conditions' (Chomsky, 2000a, p. 20). That is to say, a system that consists of mechanisms motivated entirely by the requirement that it links meanings and sounds can be called 'perfect'.

CHOMSKY'S QUESTIONS FOR LINGUISTICS

(i) *What constitutes knowledge of language?*
(ii) *How is such knowledge acquired?*
(iii) *How is such knowledge put to use?*
(iv) *What are the physical mechanisms that serve as the material basis for this system of knowledge and for the use of this knowledge?*

1.5 General ideas of language

Let us now put all this in the context of different approaches to linguistics. Chomsky's work distinguishes **externalized (E-)language** from **internalized (I-)language** (Chomsky, 1986a; 1991b). E-language linguistics, chiefly familiar from the American structuralist tradition such as Bloomfield (1933), aims to collect samples of language and then to describe their properties. An E-language approach collects sentences 'understood independently of the properties of the mind' (Chomsky, 1986a, p. 20); E-language research constructs a grammar to describe the regularities found in such a sample; 'a grammar is a collection of descriptive statements concerning the E-language' (p. 20), say that English questions involve moving auxiliaries or copulas to the beginning of the sentence. The linguist's task is to bring order to the set of external facts that make up the language. The resulting grammar is described in terms of properties of such data through 'structures' or 'patterns', as in say Pattern Grammar (Hunston and Francis, 2000).

I-language linguistics, however, is concerned with what a speaker *knows* about language and where this knowledge comes from; it treats language as an internal property of the human mind rather than something external: language is 'a system represented in the mind/brain of a particular individual' (Chomsky, 1988, p. 36). Chomsky's first question for linguistics – discovering what constitutes language knowledge – sets an I-language goal.

Chomsky claims that the history of generative linguistics shows a shift from an E-language to an I-language approach; 'the shift of focus from the dubious concept of E-language to the significant notion of I-language was a crucial step in early generative grammar' (Chomsky, 1991b, p. 10). I-language research aims to represent this mental state; a grammar describes the speaker's knowledge of the language, not the sentences they have produced. Success is measured by how well the grammar captures and explains language knowledge in terms of properties of the human mind. Chomsky's theories thus fall within the I-language tradition; they aim at exploring the mind rather than the environment. 'Linguistics is the study of I-languages, and the basis for attaining this knowledge' (Chomsky, 1987). Indeed Chomsky is extremely dismissive of the E-language approaches: 'E-language, if it exists at all, is derivative, remote from mechanisms and of no particular empirical significance, perhaps none at all' (Chomsky, 1991b, p. 10).

E-LANGUAGE AND I-LANGUAGE

E-language	**I-language**
– 'externalized'	– 'internalized'
– consists of a set of sentences	– consists of a system of principles
– deals with sentences actually produced – 'corpora'	– deals with knowledge of potential sentences – 'intuitions'
– describes properties of such data	– describes the system in an individual's mind
– is concerned with what people have done	– is concerned with what they could do

The E-language approach includes not only theories that emphasize the physical manifestations of language but also those that treat language as a social phenomenon, 'as a collection (or system) of actions or behaviours of some sort' (Chomsky, 1986a, p. 20). The study of E-language relates a sentence to the sentences that preceded it, to the situation at the moment of speaking, and to the social relationship between the speaker and the listener. It concentrates on social behaviour between people rather than on the inner psychological world. Much work within the fields of sociolinguistics and discourse analysis comes within an E-language approach in that it concerns social rather than mental phenomena.

An I-language is whatever is in the speaker's mind, namely 'a computational procedure and a lexicon' (Chomsky, 2000b, p. 119). Hence the more obvious man-in-the-street idea of language is rejected: 'In the varieties of modern linguistics that concern us here, the term "language" is used quite differently to refer to an internal component of the mind/brain' (Hauser et al., 2002, p. 1570): the things that we call the English language or the Italian language are E-languages in the social world; I-languages are known by individuals. To Chomsky the study of language as part of the mind is quite distinct from the study of languages such as English or Italian. 'a language is a state of the faculty of language, an I-language, in technical usage' (Chomsky, 2005a, p. 2).

The opposition between these two approaches in linguistics has been long and acrimonious, neither side conceding the other's reality. It resembles the perennial divisions in literature between the Romantics who saw poetry as an expression of individuality and the Classicists who saw it as part of society or, indeed, the personality difference between introverts concentrating on the world within and extraverts engaging in the world outside. It has also affected the other disciplines related to linguistics. The study of language acquisition is broadly speaking divided between those who look at external interaction and communicative function and those who look for internal rules and principles; computational linguists roughly divide into those who analyse large stretches of text and those who write rules. An E-linguist collects samples of actual speech or actual behaviour; evidence is concrete physical manifestation: what someone said or wrote to someone else in a particular time and place, say the observation that in March 2006 a Microsoft advertisement claims *Children dream of flight, to soar* or that a crossword clue in

the *Guardian* newspaper is *Sleeping accommodation with meals* or any other sentence taken at random from the billions occurring every day: if something happens, i.e. someone produces a sentence, an E-linguist has to account for it. An I-linguist invents possible and impossible sentences; evidence is whether speakers know if the sentences are grammatical – do you accept *That John left early seemed* is English or not? The E-linguist despises the I-linguist for not looking at 'real' facts; the I-linguist derides the E-linguist for looking at trivia. The I-language versus E-language distinction is as much a difference of research methods and of admissible evidence as it is of long-term goals.

The influential distinction between **competence** and **performance**, first drawn in Chomsky (1965), partly corresponds to the I-language versus E-language split. Competence is 'the speaker/hearer's knowledge of his language', performance 'the actual use of language in concrete situations' (Chomsky, 1965, p. 4). Since it was first proposed, this distinction has been the subject of controversy between those who see it as a necessary idealization and those who believe it abandons the central data of linguistics.

Let us start with Chomsky's definition of competence: 'By "grammatical competence" I mean the cognitive state that encompasses all those aspects of form and meaning and their relation, including underlying structures that enter into that relation, which are properly assigned to the specific subsystem of the human mind that relates representations of form and meaning' (Chomsky, 1980a, p. 59). The grammar of competence describes I-language in the mind, distinct from the use of language, which depends upon the context of situation, the intentions of the participants and other factors. Competence is independent of situation. It represents what the speaker knows in the abstract, just as people may know the Highway Code or the rules of arithmetic independently of whether they can drive a car or add up a column of figures. Thus it is part of the competence of all speakers of English that movements must be local. The description of linguistic competence then provides the answer to the question of what constitutes knowledge of language.

To do this, it idealizes away from the individual speaker to the abstract knowledge that all speakers possess. So differences of dialect are not its concern – whether an English speaker comes from Los Angeles or Glasgow. Nor are differences between genres – whether the person is addressing a congregation in church or a child on a swing. Nor is level of ability in the language – whether the person is a poet or an illiterate. Nor are speakers who know more than one language, say Japanese in addition to English. This is summed up in the classic *Aspects* definition of competence (Chomsky, 1965, p. 3): 'Linguistic theory is concerned primarily with an ideal speaker-listener in a completely homogenous speech community.' Ever since it was first mooted, people have been objecting that this is throwing the baby out with the bathwater; the crucial aspects of language are not this abstract knowledge but the variations and complexities that competence deliberately ignores. For example chapter 6 will argue that most human beings in fact know more than one language and that the theory cannot be arbitrarily confined to people who know only one.

Chomsky's notion of competence has sometimes been attacked for failing to deal with how language is used, and the concept of communicative competence

has been proposed to remedy this lack (Hymes, 1972). The distinction between competence and performance does not deny that a theory of use complements a theory of knowledge; I-language linguistics happens to be more interested in the theory of what people know; it claims that establishing knowledge itself logically precedes studying how people acquire and use that knowledge. Chomsky accepts that language is used purposefully; 'Surely there are significant connections between structure and function; this is not and has never been in doubt' (Chomsky, 1976, p. 56). As well as knowing the structure of language, we have to know how to use it. There is little point in knowing the structure of:

(13) Can you lift that box?

if you can't decide whether the speaker wants to discover how strong you are (a question) or wants you to move the box (a request).

Indeed in later writings Chomsky has introduced the term **pragmatic competence** – knowledge of how language is related to the situation in which it is used. Pragmatic competence 'places language in the institutional setting of its use, relating intentions and purposes to the linguistic means at hand' (Chomsky, 1980a, p. 225). It may be possible to have grammatical competence without pragmatic competence. A schoolboy in a Tom Sharpe novel *Vintage Stuff* (Sharpe, 1982) takes everything that is said literally; when asked to turn over a new leaf, he digs up the headmaster's camellias. But knowledge of language use is different from knowledge of language itself; pragmatic competence is not linguistic competence. The description of grammatical competence explains how the speaker knows that:

(14) Why are you making such a noise?

is a possible sentence of English, but that:

(15) *Why you are making such a noise?

is not. It is the province of pragmatic competence to explain whether the speaker who says:

(16) Why are you making such a noise?

is requesting someone to stop, or is asking a genuine question out of curiosity, or is muttering a *sotto voce* comment. The sentence has a structure and a form that is known by the native speaker independently of the various ways in which it can be used: this is the responsibility of grammatical competence.

Chomsky's acceptance of a notion of pragmatic competence does not mean, however, that he agrees that the sole purpose of language is communication:

> Language can be used to transmit information but it also serves many other purposes: to establish relations among people, to express or clarify thought, for creative mental activity, to gain understanding, and so on. In my opinion there is no reason to accord privileged status to one or the other of these modes. Forced to choose, I would say something quite classical and rather empty: language serves essentially for the expression of thought. (Chomsky, 1979, p. 88)

The claim that human language is a system of communication devalues the importance of other types of communication: 'Either we must deprive the notion "communication" of all significance, or else we must reject the view that the purpose of language is communication' (Chomsky, 1980a, p. 230). Though approached from a very different tradition, this echoes the sentiments of the 'British' school of linguistics from Malinowski (1923) onward that language has many functions, only one of which is communication. Chomsky claims then that 'language is not properly regarded as a system of communication. It is a system for expressing thought' (Chomsky, 2000b, p. 76). That is to say, it is a means of using the concepts of the mind via a computational system; for some purposes it might not matter if there were no interface to sounds – we can use language to organize our thoughts; it may not matter if there is no listener other than ourselves – we can talk to ourselves, write lecture notes or keep a diary. The social uses of language are in a sense secondary: 'The use of language for communication might turn out to be a kind of epiphenomenon' (Chomsky, 2002, p. 107), an accidental side-effect. Hence question (iii) 'how is knowledge put to use?' can only be tackled after we have described the knowledge itself, question (i).

In all Chomskyan models a crucial element of competence is its creative aspect; the speaker's knowledge of language must be able to cope with sentences that it has never heard or produced before. E-language depends on history – pieces of language that happen to have been said in the past. I-language competence must deal with the speaker's ability to utter or comprehend sentences that have never been said before – to understand:

(17) Ornette Coleman's playing was quite sensational.

even if they are quite unaware who Ornette Coleman is or what is being talked about. It must also reflect the native speakers' ability to judge that:

(18) *Is John is the man who tall?

is an impossible sentence, even if they are aware who is being referred to, and can comprehend the question; 'having mastered a language, one is able to understand an indefinite number of expressions that are new to one's experience, that bear no simple physical resemblance to the expressions that constitute one's linguistic experience' (Chomsky, 1972a, p. 100), whether a sentence from a radio presenter such as:

(19) If you have been, thank you for listening.

or today's newspaper headline:

(20) Bomb death day before birthday.

Creativity in the Chomskyan sense is the mundane everyday ability to create and understand novel sentences according to the established knowledge in the mind – novelty within the constraints of the grammar. 'Creativity is predicated on a

system of rules and forms, in part determined by intrinsic human capacities. Without such constraints, we have arbitrary and random behaviour, not creative acts' (Chomsky, 1976, p. 133). It is not creativity in an artistic sense, which might well break the rules or create new rules, even if ultimately there may be some connection between them. The sentence:

(21) There's a dog in the garden.

is as creative as:

(22) There is grey in your hair.

in this sense, regardless of whether one comes from a poem and one does not.

It is then a characteristic of human language that human beings can produce an infinite number of sentences. I-language has to deal with 'the core property of discrete infinity' of language (Hauser et al., 2002, p. 1571). The ability of language to be infinite is ascribed by Chomsky to recursion, meaning that some rules of language can have other versions of themselves embedded within them indefinitely, rather like fractals. Or indeed the example in *The Mouse and his Child* (Hoban, 1967), where a can of dog-food has a picture of a dog looking at a can of dog-food with a picture of a dog . . . , and so on till the last visible dog. Recursion will be discussed in chapter 3.

Let us now come back to performance, the other side of the coin. One sense of performance corresponds to the E-language collection of sentences. In this sense performance means any data collected from speakers of the language – today's newspaper, yesterday's diary, the improvisations of a rap singer, the works of William Shakespeare, everything anybody said on TV yesterday, the 100 million words in the British National Corpus. Whether it is W. B. Yeats writing:

(23) There is grey in your hair.

or a band leader saying:

(24) Write down your e-mail address if you would like to be let know about our next CD.

it is all performance. An E-language grammar has to be faithful to a large sample of such language, as we see from recent grammars such as Biber et al. (1999), based on a corpus of 40 million words. An I-language grammar does not rely on the regularities in a collection of data; it reflects the knowledge in the speaker's mind, not their performance.

However, 'performance' is used in a second sense to contrast language knowledge with the psychological processes through which the speaker understands or produces language. Knowing the Highway Code is not the same as being able to drive along a street; while the Code in a sense informs everything the driver does, driving involves a particular set of processes and skills that are indirectly related to knowledge of the Code. Language performance has a similar relationship

to competence. Speakers have to use a variety of psychological and physical processes in actually speaking or understanding that are not part of grammatical competence, even if they have some link to it; memory capacity affects the length of sentence that can be uttered but is nothing to do with knowledge of language itself. Samples of language include many phenomena caused by these performance processes. Speakers produce accidental spoonerisms:

(25) You have hissed my mystery lectures and will have to go home on the town drain.

and hesitations and fillers such as *er* and *you know:*

(26) I, er, just wanted to say, well, you know, have a great time.

They get distracted and produce ungrammatically odd sentences:

(27) At the same time do they see the ghost?

They start the sentence all over again:

(28) Well it was the it was the same night.

One reason for the I-linguist's doubts about using samples of language as evidence for linguistic grammars is that they reflect many other psychological processes that obscure the speaker's actual knowledge of the language. The other half of Chomsky's *Aspects* definition of competence (Chomsky, 1965, p. 3) is that the ideal speaker-hearer is: 'unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance'. We need to see through the accidental properties of the actual production of speech to the knowledge lying behind it. Again, censoring all these performance aspects from competence has not been an easy concept for many people to accept: 'I believe that the great majority of psycholinguists around the world consider the competence–performance dichotomy to be fundamentally wrong-headed' (Newmeyer, 2003, p. 683).

COMPETENCE AND PERFORMANCE

Competence: 'the speaker-hearer's knowledge of his language' (Chomsky, 1965, p. 4)
Performance: 'the actual use of language in concrete situations' (Chomsky, 1965, p. 4)
Pragmatic competence: 'places language in the institutional setting of its use, relating intentions and purposes to the linguistic means at hand' (Chomsky, 1980a, p. 225)

EXERCISE 1.2 Here is an example of performance – a transcript of an ordinary recorded conversation. What aspects of it would you have to ignore to produce a version that more accurately reflected the speaker's competence?

A: After the the initial thrill of being able to afford to stay in a hotel wore off I found them terrible places. I mean perhaps if one can afford a sufficiently expensive one, it's not like that but er these sort of smaller hotels where it's like a private house but it's not your private house and you can't behave as you would at home.
B: Surely that's dying out in England now? . . .
A: London is absolutely full of them – small private hotels and you're supposed to be in at a particular time at night, let's say half past eleven, and breakfast will be served between half past seven and half past eight and if you come downstairs too late, that's it.

1.6 Linguistic universals

Let us now consider the notion of a linguistic universal a little more closely. One might think that something that is universal should be present in all linguistic systems. However, this is not necessarily the case. Consider the notion of **movement** for example. Movement plays an important role in Chomskyan theory and is employed to describe a number of constructions ranging from passives to questions. In English question words typically begin with the letters 'wh', i.e. *who, where,* and so on, and are therefore called **wh-elements**. A question may be formed by moving a wh-element to the front of the sentence. Suppose a person called *Mike* knows a person called *Bert*. This can be expressed by the sentence:

(29) Mike knows Bert.

Suppose now that, although we know that Mike knows someone, we don't know who that person is. We might then ask the question:

(30) Who does Mike know?

Note here the wh-element *who* stands for the object of the verb (the one who is known). Objects in English usually follow the verb, as in sentence (29), yet in the question (30) the wh-element occupies a position at the front of the sentence. We might therefore propose that forming the question involves moving the object wh-element from its position behind the verb to its interrogative position at the front of the sentence:

(31) Who does Mike know – ?

English wh-questions then involve movement; some element moves from its usual position to the front of the sentence.

But not every movement apparent in one language is apparent in all. In Japanese for example the statement:

(32) Niwa-wa soko desu.
(garden there is)
The garden is there.

differs from the question:

(33) Niwa-wa doko desu ka?
(garden where is)
Where is the garden?

by adding the element *ka* at the end and having the question word *doko* in the place of *soko*. The question-word *doko* is not moved to the start, as must happen in English (except for echo questions such as *You said what?*). Japanese does not use syntactic movement for questions, though it may need other types of movement.

Other languages also lack movement for questions. In Bahasa Malaysia for example they can be formed by adding the question element *kah* to the word that is being asked about (King, 1980):

(34) Dia nak pergi ke Kuala Lumpurkah?
(he is going to Kuala Lumpur)
Is he going to Kuala Lumpur?

without shifting it to the front. The presence or absence of syntactic movement is then a parameter of variation between languages; English requires certain movements, Bahasa Malaysia and Japanese do not. Parameters such as presence or absence of question movement vary in setting from one language to another. If knowledge of language were just a matter of fixed principles, all human languages would be identical; the variation between them arises from the different way that they handle certain parameterized choices, such as whether or not to form questions through movement.

The setting of this movement parameter has a chain of consequences in the grammar. For example, as will be elaborated on in the next chapter, there are restrictions on movements which hold for all languages. One of these is that movements have to be short, a result of a principle (or principles) known as the Locality Conditions. But clearly a language which does not have wh-movement has no need for the locality conditions placed on such movements since there is nothing for them to affect.

In what sense can a universal that does not occur in every language still be universal? Japanese does not *break* any of the requirements of syntactic movement; it does not need locality for question movement because question movement itself does not occur. Its absence from some aspect of a given language does not prove it is not universal. Provided that the universal is found in some human language, it does not have to be present in all languages. UG Theory does not insist all languages are the same; the variation introduced through parameters allows principles to be all but undetectable in particular languages. UG Theory does not, however, allow principles to be *broken*.

This approach to universals can be contrasted with the longstanding attempts to construct a typology for the languages of the world by enumerating what they have in common, leading to what are often called 'Greenbergian' universals, after the linguist Joseph Greenberg. One example is the Accessibility Hierarchy (Keenan and Comrie, 1977). All languages have relative clauses in which the subject of the relative clause is related to the Noun as in the English:

(35) Alexander Fleming was the man who discovered penicillin.

A few languages do not permit relative clauses in which the object in the relative clause relates to the Noun. For example the English:

(36) This is the house that Jack built.

would not be permitted in Malagasy. Still more languages do not allow the indirect object from the relative clause to relate to the Noun. The English sentence:

(37) John was the man they gave the prize to.

would be impossible in Welsh. Further languages cannot have relative clauses that relate to the Noun via a preposition, as in:

(38) They stopped the car from which the number plate was missing.

or via a possessive; the English sentence:

(39) He's the man whose picture was in the papers.

would not be possible in Basque. Unlike many languages, English even permits, though with some reluctance, the relative clause to relate via the object of Comparison as in:

(40) The building that Canary Wharf is taller than is St Paul's.

The Accessibility Hierarchy is represented in terms of a series of positions for relativization:

(41) Subject > Object > Indirect Object > Object of Preposition > Genitive > Object of Comparison.

All languages start at the left of the hierarchy and have subject relative clauses; some go one step along and have object clauses as well; others go further along and have indirect objects; some go all the way and have every type of relative clause, including objects of comparison. It is claimed that no language can avoid this sequence; a language may not have, say, subject relative clauses and object of preposition relative clauses but miss out the intervening object and indirect object clauses. The Accessibility Hierarchy was established by observations

based on many languages; it is a Greenbergian universal. There is as yet no compelling reason within UG Theory why this should be the case, no particular principle or parameter involved: it is simply the way languages turn out to be. Greenbergian universals such as the Accessibility Hierarchy are data-driven; they arise out of observations; a single language that was an exception could be their downfall, say one that had object of preposition relative clauses but no object relative clauses. Universals within UG Theory are theory-driven; they may not be breached but they need not be present. There may indeed be a UG explanation for a particular data-driven universal such as the Accessibility Hierarchy; this would still not vitiate the distinction between the theory-driven UG type of universal and the data-driven Greenbergian type.

So it is not necessary for a universal principle to occur in dozens of languages. UG research often starts from a property of a single language, such as locality in English. If the principle can be ascribed to the language faculty itself rather than to experience of learning a particular language, using an argument to be developed in chapter 4, it can be claimed to be universal on evidence from one language alone; 'I have not hesitated to propose a general principle of linguistic structure on the basis of observations of a single language' (Chomsky, 1980b, p. 48). Newton's theory of gravity may have been triggered by an apple but it did not require examination of all the other apples in the world to prove it. Aspects of UG Theory are disprovable; a principle may be attributed to UG that further research will show is in fact peculiar to Chinese or to English; tomorrow someone may discover a language that clearly breaches locality. The nature of any scientific theory is that it can be shown to be wrong; 'in science you can accumulate evidence that makes certain hypotheses seem reasonable, and that is all you can do – otherwise you are doing mathematics' (Chomsky, 1980b, p. 80). Locality is a current hypothesis, like gravity or quarks; any piece of relevant evidence from one language or many languages may disconfirm it. But of course, the evidence does have to be relevant: not just the odd observation that seems to contradict the theory but a coherent alternative analysis. Producing a single sentence that contradicts a syntactic theory, say, is not helpful if we cannot also provide an account for the problematic sentence. There are anomalous facts that do not fit the standard theory in most fields, whether physics or cookery, but they are no particular use until they lead to a broader explanation that subsumes them and the facts accounted for by the old theory.

UNIVERSALS

- Universals consist of principles such as locality.
- Chomskyan universals do not have to occur in all languages, unlike Greenbergian universals.
- No language violates a universal principle (the language simply may not use the principle in a particular context).
- Universals are part of the innate structure of the human mind and so do not have to be learnt (to be discussed later).

1.7 The evidence for Universal Grammar Theory

How can evidence be provided to support these ideas? The question of evidence is sometimes expressed in terms of 'psychological reality'. Language knowledge is part of the speaker's mind; hence the discipline that studies it is part of psychology. Chomsky has indeed referred to 'that branch of human psychology known as linguistics' (Chomsky, 1972a, p. 88). Again, it is necessary to forestall too literal interpretations of such remarks. A movement rule such as:

(42) Move a wh-element to the front of the clause.

does not have any necessary relationship to the performance processes by which people produce and comprehend sentences; it is solely a description of language knowledge. To borrow a distinction from computer languages, the description of knowledge is 'declarative' in that it consists of static relationships, not 'procedural' in that it does not consist of procedures for actually producing or comprehending speech. The description of language in rules such as these may at best bear some superficial resemblance to speech processes.

When knowledge of language is expressed as principles and parameters, the resemblance seems even more far-fetched; it is doubtful whether every time speakers want to produce a sentence they consider in some way the interaction of all the principles and parameters that affect it. Conventional psychological experiments with syntax tell us about how people perform language tasks but nothing directly about knowledge. They can provide useful indirect confirmation of something the linguist already suspects and so give extra plausibility perhaps. But they have no priority of status; the theory will not be accepted or rejected as a model of knowledge because of such evidence alone. Chomsky insists that the relevant question is not 'Is this psychologically real?' but 'Is this true?' He sees no point in dividing evidence up into arbitrary categories; 'some is labelled "evidence for psychological reality", and some merely counts as evidence for a good theory. Surely this position makes absolutely no sense . . . ?' (Chomsky, 1980a, p. 108). The linguist searches for evidence of locality and finds that questions such as:

(43) Is John the man who is tall?

are possible, but questions such as:

(44) *Is John is the man who tall?

are impossible. The linguist may then look for other types of confirming evidence: how speakers form questions in an experiment, the sequence in which children acquire types of question, the kinds of mistake people make in forming questions. All such evidence is grist to the mill in establishing whether the theory is correct; 'it is always necessary to evaluate the import of experimental data on theoretical constructions, and in particular, to determine how such data bear on hypotheses

that in nontrivial cases involve various idealizations and abstractions' (Chomsky, 1980b, p. 51). Psychological experiments provide one kind of data, which is no more important than any other. A speaker's claim that:

(45) Is John the man who is tall?

is a sentence of English and:

(46) *Is John is the man who tall?

is not provides a concrete piece of evidence about language knowledge. It doesn't matter whether the sentence has ever actually been said, only whether a sentence of that form *could* be said and how it would be treated if it *were*. Such evidence may well be incomplete and biased; in due course it will have to be supplemented with other evidence. But we have to start somewhere. The analysis of this easily available evidence is rich enough to occupy generations of linguists. It is simplest to start from the bird in the hand, our own introspections into the knowledge of language we possess; when that has been exhausted, other sources can be tapped.

However, it should not be thought that the UG Theory attempts to deal with everything about language. UG Theory recognizes that some aspects of language are unconnected to UG. For example in English the usual form of the past tense morpheme is '-ed', pronounced in three different ways as /d/ *planned*, /t/ *liked*, and /ɪd/ *waited*. But English also has a range of irregular past forms, some that form definite groups such as vowel change *bite/bit*, or lack of change *hit/hit*, others that have no pattern *go/went*, *am/was*. These irregular forms are learnt late by children in the first language, give problems to learners of a second language, and vary from one region to another, UK *dived* versus US *dove*. There is no need for UG to explain this odd assortment of forms; they are simply facts of English that any English speaker has to learn, unconnected to UG. The fact that such forms could be learnt in associationist networks has been used as a test case by advocates of connectionism to show that 'rules' are not needed (Rumelhart and McLelland, 1986) (though this often involves a confusion of the spoken and written language – /peɪd/ is regular in speech, <paid> irregular in writing (Cook, 2004)). As these forms are marginal for UG, whether they are learnable or not by such means has no relevance to the claims of the UG Theory.

The knowledge of any speaker contains masses of similar oddities that do not fit the overall pattern of their language. 'What a particular person has in the mind/brain is a kind of artefact resulting from the interplay of accidental features' (Chomsky, 1986a, p. 147). UG Theory avoids taking them all on board by raising a distinction between core and periphery. The core is the part of grammatical competence covered by UG where all the principles are maintained, all the parameters set within the right bounds. The periphery includes aspects that are not predictable from UG. It is unrealistic to expect UG Theory to account for myriads of unconnected features of language knowledge. It deals instead with a core of central language information and a periphery of less essential information; 'a core language is a system determined by fixing values for the parameters

of UG, and the periphery is whatever is added on in the system actually represented in the mind/brain of a speaker-hearer' (Chomsky, 1986a, p. 147). The theory of UG is far from a complete account of the speaker's entire knowledge of language; it deals with the core aspects that are related to UG, not with the periphery that is unrelated to UG.

EXERCISE 1.3 Here are some sentences of Gateshead English, as represented in a novel. Setting aside differences in accent, try to allocate them to core grammar, to performance or to peripheral grammar.

1 'How we gonna gan and see them?'
2 'Ah cannet wait, man Gerry.'
3 'Ah've always wanted a dog, me.'
4 'Cannet hang around here all night.'
5 'When was the last time ye went to a match like?'
6 'Neeone's gonnae burn doon nee hoose.'
7 'Me mate Sewell here doesn't like being conned.'
8 'Bet even Gazza couldn't get a penalty past Rusty.'

Source: Jonathan Tulloch (2000), *The Season Ticket*, Jonathan Cape

---

1.8 Conclusion

To sum up, the distinctive feature of Chomsky's I-language approach is that its claims are not unverifiable but checkable; it supports unseen mental ideas with concrete evidence. The theory can easily be misconceived as making woolly abstract statements unconnected to evidence, which can be countered by sheer assertion and argument. Most criticism of Chomskyan concepts has indeed attempted to refute them by logic and argument rather than by attacking their basis in data and evidence.

A case in point is Chomsky's well-known argument that children are born equipped with certain aspects of language, based on the fact that children know things about language that they could not have learnt from the language they have heard, as developed in chapters 2 and 5. This argument could be refuted by showing that the alleged fact is incorrect: either children could learn everything about language from what they hear or adults do not have the knowledge ascribed to them. 'An innatist hypothesis is a refutable hypothesis' (Chomsky, 1980a, p. 80). But the argument cannot be dismissed by pure counter-argument unsupported by evidence.

The discussion in this book describes the actual syntactic content of the theory at length because Chomsky's general ideas cannot be adequately understood without looking at the specific claims about language on which they are based. A principle of language is not a proposal for a vague abstraction but a specific hypothesis about the facts of human language, eventually coming down to precise claims about the grammaticality or ungrammaticality of particular sentences.

The UG Theory claims to be a scientific theory based on solid evidence about language. As such, it is always progressing towards better explanations for language knowledge, as later chapters will demonstrate.

Discussion topics

1 To what extent do you think Chomsky's theories as presented in this chapter are in fact scientific?
2 Is describing the core of language as the computational system a change of label or something more profound?
3 Can you accept Chomsky's claim that not all universal principles will be found in every language?
4 Do Chomsky's questions for linguistics touch on any issues that the person in the street would recognize?
5 Is it really creative to be able to produce new sentences? Couldn't a computer do the same?
6 What do you think could count as evidence against Chomsky's theories?
7 How acceptable do you find a scientific approach that can dismiss any counter-examples as peripheral grammar or performance? Or is it inevitable that any scientific theory has to be able to discard some of its apparent data?
8 For those who know Chomsky's political views in books such as Chomsky (2004a), what connection do you see between them and his overall ideas of language?

2 Principles, Parameters and Language Acquisition

This chapter takes a closer look at some of the ideas that have been driving Chomskyan linguistics for more than twenty-five years. It explains the notions of principles and parameters, introduced briefly in the last chapter, and shows how these are integrated with ideas about language acquisition.

2.1 Principles and parameters

2.1.1 Rules in early generative grammar

To understand grammatical principles and parameters means looking at certain linguistic phenomena that they account for and sketching what these notions replaced.

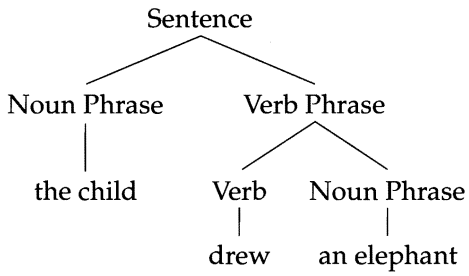
2.1.2 Phrase structure and rewrite rules

A major assumption in linguistics since the 1930s has been that sentences consist of phrases – structural groupings of words: sentences have **phrase structure**. Thus the **sentence (S)**:

(1) The child drew an elephant.

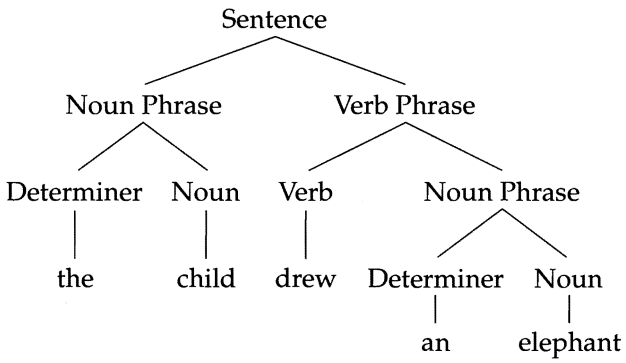
breaks up into a **Noun Phrase (NP)** *the child* and a **Verb Phrase (VP)** *drew an elephant*. The VP in turn breaks up into a **Verb (V)** *drew* and a further Noun Phrase *an elephant*:

(2)



These Noun Phrases also break up into smaller constituents; the NP *the child* consists of a **Determiner (Det** or **D)** *the* and a **Noun (N)** *child*, while the NP *an elephant* consists of a Determiner *an* and an N *elephant*. The final constituents are then items in the lexicon:

(3)



Phrase structure analysis thus breaks the sentence up into smaller and smaller grammatical constituents, finishing with words or morphemes when the process can go no further. A sentence is not just a string of words in a linear sequence but is structured into phrases, all of which connect together to make up the whole. A sentence is then defined by the phrases and lexical items into which it expands.

A **tree diagram** such as (3) is one way to represent the phrase structure of a sentence: each constituent of the structure is represented by a **node** on the tree, which is labelled with its name; elements which are grouped into a constituent are linked to the node by **branches**. Another way of representing structure commonly found in linguistic texts is through labelled brackets, where paired brackets are used to enclose the elements that make up a constituent and the label on the first of the pair names the constituent. Thus the structure in (3) might equally be represented as the following labelled bracketing without any change in the analysis:

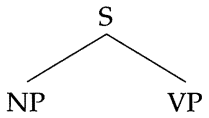
(4) $[_{S}\,[_{NP}\,[_{D}\,\text{The}]\ [_{N}\,\text{child}]][_{VP}\,[_{V}\,\text{drew}][_{NP}\,[_{D}\,\text{an}]\ [_{N}\,\text{elephant}]]]]$.

One of Chomsky's first influential innovations in linguistics was a form of representation for phrase structure called a **rewrite rule** (Chomsky, 1957), seen in:

(5) $S \rightarrow NP\ VP$

In this the 'rewrite' arrow '→' can be taken to mean 'consists of'. The rule means exactly the same as the phrase structure tree:

(6)



or as the bracketing:

(7) $[_S$ NP VP]

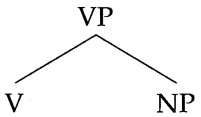
namely that the Sentence (S) 'consists of' a Noun Phrase (NP) and a Verb Phrase (VP).

The next rewrite rule for English will then be:

(8) VP → V NP

Again, this means that the VP 'consists of' a Verb (V) and an NP, as shown in the tree:

(9)



or as the bracketing:

(10) $[_{VP}$ V NP]]

More rewrite rules like (5) and (8) can be added to include more and more of the structure of English, for example to bring in the structure of Noun Phrases:

(11) NP → Det N

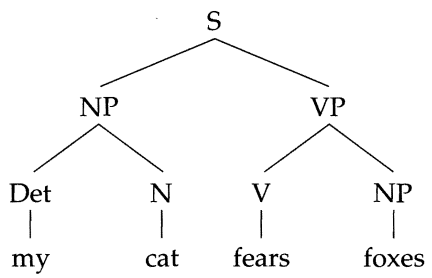
showing that NPs consist of a determiner (Det) and a noun (N). The sentence is generated through the rewrite rules until only lexical categories like Noun or Verb are left that cannot be further rewritten. At this point an actual word is chosen to fit each category out of all the appropriate words stored in the lexicon, say *child* for N and *draw* for V.

Rewrite rules, then, take the subjective element out of traditional grammar rules. Understanding a rule like 'Sentences have subjects and predicates' means interpreting what you mean by a subject or predicate. To understand rule (5) you don't need to know what an S is because the very rule defines it as an NP and a VP; you don't need to know what an NP is because rule (11) defines it as a Determiner and a Noun; and you don't need to know what a Noun is because you can look up a list or a dictionary. Rewrite rules are formal and explicit, sometimes called 'generative' as we see below. Hence they can easily be used by

computers. Rewrite rules can be written directly in to the computer language Prolog, so that you can type in a set of rules like (5) and get a usable parser for generating English sentence structures (Gazdar and Mellish, 1989).

PHRASE STRUCTURE

Phrase structure analysis divides sentences into smaller and smaller constituents until only words or morphemes are left, usually splitting into two constituents at each point, most commonly represented as a tree diagram:



or as labelled brackets:

$[_{S}\ [_{NP}\ [_{Det}$My cat$][_{VP}$ fears $_{NP}$[foxes]]]].

Rewrite rules formally define a set of possible structures which model those of any particular language, thus making a grammar of that language. We might therefore expect natural languages to be made up of such rewrite rules. However, the rules given here are specific to particular structures in particular languages. So, for example, rule (5) S → NP VP does not necessarily imply that *all* sentences of *all* languages consist of an NP subject followed by a VP. In some languages the subject comes last, as in Malagasy:

(12) Nahita ny mpianatra my vehivavy.
(saw the student the woman)
The woman saw the student.

thus requiring another rewrite rule:

(13) S → VP NP

Even in English, few sentences have the bare NP V NP structure seen in (3). They may for instance contain other elements, such as adverbs *often* or auxiliary verbs *will*. The NPs and VPs may differ from the ones seen so far by containing adjectives *old* or having intransitive verbs with no objects *runs*. To work for English, the rewrite rules in (3), (8) and (11) need to be massively expanded in number. Malagasy would similarly need vast expansion from the single rule given in (13).

The grammars of the two languages have to be spelled out specifically in rewrite rules and have very little in common. The grammars of languages differ in terms of the actual rewrite rules even if they can be described in the same formal way. Rewrite rules are not specific to a language, so they can never achieve the overall goal of universality for all languages.

A MINI-GRAMMAR FOR ENGLISH USING REWRITE RULES

S → NP VP
VP → V NP
NP → Det N

Lexicon
Ns: child, elephant, truth, Winston Churchill . . .
Vs: draw, fly, cook, pay, invigilate . . .
Dets: the, a, an . . .

EXERCISE 2.1

1 In the early days of rewrite rules, people tried to use them for a variety of other purposes such as classification of cattle-brands (Watt, 1967). Try to write a system of rewrite rules like the one in the box above for one or more of the following:

- a three course meal
- building a house
- yesterday's television programmes
- the structure of *The Lord of the Rings*, or another novel
- dancing the salsa
- making a cup of tea

2 Below is a description of the English sentence

The policeman arrested the man.

in traditional grammatical terms. Can you convert it into a tree diagram? Into rewrite rules?

English sentences have a subject and a predicate; predicates may have transitive verbs with an object; singular nouns may have a definite article.

2.1.3 *Movement*

Phrase structure rewrite rules are not enough for natural languages. A major assumption in most linguistic theories, harking back to traditional grammars, is

that elements of the sentence appear to move about, as we discussed in the last chapter. A sentence such as:

(14) What did the child draw?

makes us link the question-word *what* to the element in the sentence that is being questioned, namely the object:

(15) The child drew something. (What was that something?)

as if *what* had been moved to the beginning of the sentence from its original position after the verb *drew*:

(16) The child drew what.

What did the child draw ?

This can be termed the 'displacement property' of language: 'phrases are interpreted as if they were in a different position in the expression' (Chomsky, 2000b, p. 12). The relationship between the displaced elements and the original position is called 'movement': a phrase is said to 'move' from one position in the structure to another. This does not mean *actual* movement as we process the sentence, i.e. the speaker thinking of an element and then mentally moving it, even if this mistaken implication at one stage led to a generation of psycholinguistic research; movement is an abstract relationship between two forms of the sentence, which behave *as if* something moved. In other words movement is a relationship in competence – the knowledge of language – not a process of performance.

The kind of rule that deals with displacement phenomena is called a **transformation**. In general such rules take a structure generated by the rewrite rules and transform it into a different structure in which the various constituents occupy different positions. Again, such rules are specific to a particular language and specific to particular constructions rather than applying to all constructions in all languages. Other English constructions involving displacements include the passive:

(17) The elephant was drawn.

In this sentence the NP in subject position, *the elephant*, is interpreted as the object and hence we might propose that it involves a movement from object to subject position (with some other changes):

(18) drew the elephant

This is clearly very different from questions where a wh-element moves to a position in front of the subject, not to the subject position itself.

We might capture these facts by claiming that English grammar consists of many such transformational rules, one for each construction that involves displacement. Hence these rules are specific to the constructions that they are involved in.

Furthermore, while English has a transformation rule which moves an interrogative word like *what* to the beginning of a question-sentence, Japanese does not demonstrate this displacement:

(19) a. Niwa-wa soko desu.
(garden there is)
The garden is there.

b. Niwa-wa soko desu ka?
(garden there is Q)
Where is the garden?

As can be seen from the examples in (19), a Japanese question is formed by the addition of an interrogative particle *ka* at the end of the sentence; the other elements remain in the same positions they occupy in the declarative. On the other hand, movements occur in other languages which do not happen in English. Hungarian, for example, moves a focused element in front of the verb where English would use stress or a particular sentence type, known as a cleft sentence (seen in the English translation in (20), to express the same thing:

(20) János tegnap ment el.
(John yesterday left)
John left YESTERDAY.
It was yesterday that John left.

Since languages use different transformation rules, their grammars must obviously differ in terms of the transformation rules they contain.

EXERCISE 2.2 How could you treat the following sentences as examples of movement in English? What has moved and where?

Where have all the flowers gone?
Do you think I'm sexy?
What's new pussycat?
How many roads must a man walk down?
Why, why, why Delilah?
Are you lonely tonight?
When the saints go marching in.
In the town where I was born lived a man who sailed to sea.
Is that all there is?

The picture of grammar that emerges from these observations is a collection of construction-specific phrase structure and transformational rules which differ from language to language. This view prevailed in generative circles throughout the 1960s and 1970s, until it was challenged by Principles and Parameters (P&P) Theory in the 1980s, which claimed that natural language grammars are not constructed

out of these kinds of rules, but of something far more general. Principles state conditions on grammaticality that are not confined to a specific construction in a specific language, but are in principle applicable to *all* constructions in *all* languages. We will exemplify a possible principle later in the chapter.

2.1.4 *A brief word on the meaning of 'generative'*

Before going further into principles and parameters, it is worth considering the notion of *generative* used in generative grammar, first to avoid a common misunderstanding and secondly to see what has happened to the term since the onset of the P&P approach. 'Generative' means that the description of a language given by a linguist's grammar is rigorous and explicit: 'when we speak of the linguist's grammar as a "generative grammar" we mean only that it is sufficiently explicit to determine how sentences of the language are in fact characterised by the grammar' (Chomsky, 1980a, p. 220). The chief contrast between traditional grammar statements and the rules of generative grammar lay not in their content so much as in their expression; generative rules are precise and testable without making implicit demands on the reader's knowledge of the language. One of the famous traps people fall into, called the Generative Gaffe by Botha (1989), is to use the term 'generative' as a synonym for 'productive' rather than for 'explicit and formal'. A rewrite rule like (5) is intended as a model not directly for how people produce sentences but for what they know. Thus a generative grammar is not like an electrical generator which 'generates' (i.e. produces) an electrical current: when we say that a grammar generates a language, we mean that it describes the language in an explicit way. A set of phrase structure and transformational rules form a generative grammar, as they state precisely what the structures are in a language and how those structures may be transformed into other structures. This contrasts with traditional grammars which might tell us, for example, that 'the subject comes before the verb' without defining what 'the subject' is or stating explicitly where 'before the verb' is.

However, the replacement of rules by principles has consequences for the interpretation of the term 'generative'. Principles do not readily lend themselves to the same formal treatment as rules. The rival syntactic theory of Head-Driven Phrase Structure Grammar (HPSG) (Pollard and Sag, 1994) claims firmly to be part of generative grammar on the grounds that it uses formal explicit forms of statement, but challenges the right of recent Chomskyan theories to be called 'generative'; generative grammar 'includes little of the research done under the rubric of the "Government Binding" framework, since there are few signs of any commitment to the explicit specification of grammars or theoretical principles in this genre of linguistics' (Gazdar et al., 1985, p. 6). One introduction to 'generative grammar' (Horrocks, 1987) devoted about half its pages to Chomskyan theories, while a survey chapter on 'generative grammar' (Gazdar, 1987) dismissed Chomsky in the first two pages. Thus, though the Universal Grammar (UG) Theory still insists that grammar has to be stated explicitly, this is no longer embodied in the formulation of actual rules. The rigour comes in the principles and in the links to evidence.

Although Chomsky later claimed 'true formalization is rarely a useful device in linguistics' (Chomsky, 1987), the theory nevertheless insists on its scientific status as a generative theory to be tested by concrete evidence about language. It sees the weakness of much linguistic research as its dependence on a single source of data – observations of actual speech – when many other sources can be found. A scientific theory cannot exclude certain things in advance; the case should not be prejudged by admitting only certain kinds of evidence. 'In principle, evidence . . . could come from many different sources apart from judgments concerning the form and meaning of expression: perceptual experiments, the study of acquisition and deficit or of partially invented languages such as Creoles, or of literary usage or language change, neurology, biochemistry, and so on' (Chomsky, 1986a, pp. 36–7).

Some of these disciplines may not as yet be in a position to give hard evidence; neurology may not be able to show clearly how or where language is stored physically in the brain: in principle it has relevant evidence to contribute and may indeed do so one day. Fodor (1981) contrasts what he calls the 'Wrong View' that linguistics should confine itself to a certain set of facts with the 'Right View' that 'any facts about the use of language, and about how it is learnt . . . could in principle be relevant to the choice between competing theories'. When UG Theory is attacked for relying on intuitions and isolated sentences rather than concrete examples of language use or psycholinguistic experiments, its answer is to go on the offensive by saying that in principle a scientific theory should not predetermine what facts it deals with.

The term 'generative grammar' is, however, now generally used by many linguists simply as a label for research within the UG paradigm of research, for example the conference organizations Generative Linguists of the Old World (GLOW) and Generative Approaches to Second Language Acquisition (GASLA). It is often now little more than a vaguely worthy term distinguishing some Chomsky-derived linguistics approaches from others rather than having precise technical content.

2.1.5 *An example of a principle: locality*

To show what is meant by a grammatical principle we can develop the idea of movement. The previous section introduced the view that certain elements move about in a structure. As might be expected, movement has limits; many possible movements result in a structure that is ungrammatical. One of the first limitations to be noticed was that movements have to be short, i.e. not span too much of the sentence. Much work on movement since the late 1960s has tried to account for this observation by proposing different principles, which crop up in subsequent chapters. For now, however, this limitation can be called the **Locality Principle**: movements must be within a 'local' part of the sentence from which the moved element originates.

Some examples will serve to put the idea across. To form a yes-no question in English (i.e. one that can be answered with either a 'yes' or a 'no') an auxiliary

verb, such as *will, can, may, have* or *be,* moves from its normal position behind the subject to a position in front of the subject:

(21) The manager will fire Beckham.

(22) Will the manager fire Beckham?

This is known as **subject–auxiliary inversion**. This example is simple as there is just one auxiliary verb which moves to form the question. However, other sentences have more than one auxiliary verb:

(23) The manager will have fired Beckham.

The issue is whether either of the two auxiliaries *will* and *have* may move to form the question:

(24) a. Will the manager have fired Beckham?
b. * Have the manager will fired Beckham?

In this case the first auxiliary can move, but the second cannot, due to the Locality Principle. Comparing the distances that the two auxiliaries must move in these examples, moving the first auxiliary *will* clearly involves a shorter movement than moving the other auxiliary *have*:

(25) a. Will the manager – have fired Beckham?
b. Have the manager will – fired Beckham?

In other words, the shorter movement is grammatical and the longer ungrammatical.

This observation applies not only to the movement of auxiliary verbs but also to other movements. The following example involves moving a subject of an embedded clause *the manager* to a higher subject position in the sentence:

(26) a. It seems [the manager has fired Beckham].
b. The manager seems [– to have fired Beckham].

This is known as **subject raising**. If there is more than one embedded clause, movement from one subject position to the next is possible, but not movement over the top of a subject:

(27) a. It seems [the manager is likely [– to fire Beckham]].
b. * The manager seems [it is likely [– to fire Beckham]].

Again, the shorter of the two movements is grammatical, the longer one is ungrammatical, conforming to the Locality Principle.

Let us consider another type of movement. As discussed in the previous section, certain questions are formed by displacing question words, known as **wh-elements**, to the front of the sentence:

(28) Who did the manager fire – ?

This is called **wh-movement**. Although the issue will need more discussion in chapter 4, wh-movements also have to be short. As seen in (28), a wh-element is allowed to move out of a clause to a position at the front of the clause. But moving another wh-element to the front of an even higher clause gives an ungrammatical result:

(29) a. David asked [who the manager fired –].
b. * Who did David ask [who – fired –]?

Thus, while a wh-element can move to the front of its own clause, it cannot move directly to the front of a higher clause, which would obviously involve a longer movement. Again, wh-movement obeys the Locality Principle.

The Locality Principle extends to other linguistic phenomena as well as movements. Take the case of English pronouns. These are said to 'refer to' nouns in the sentence. In:

(30) Peter met John and gave him the message.

the personal pronoun *him* refers to the same person as *John*. Reflexive pronouns such as *myself* and *herself*, however, obey strict conditions on what they can refer to, which differ from those for personal pronouns such as *me* and *her*. For one thing, a reflexive pronoun must refer to some other element in the same actual sentence and cannot refer to something directly identified in the discourse situation outside the sentence, as personal pronouns can:

(31) a. George talks to himself.
b. * George talks to herself.
c. George talks to her.

In (31a) the reflexive pronoun *himself* obviously refers to *George* and no one else. (31b), however, is ungrammatical as the reflexive *herself* cannot refer to anything in the sentence and, unlike the personal pronoun *her* in (31c), it cannot refer to someone not mentioned in the sentence either, with the minor well-known exception of George Eliot, the Victorian woman novelist. However, if there are two possible antecedents, i.e. elements that a pronoun can refer to, one inside the clause containing the reflexive and one outside this clause, only the nearest one can actually be referred to:

(32) The regent thinks [George talks to himself].

In this sentence only *George* can act as the antecedent of *himself*, although in other sentences *the regent* would be a possible antecedent. The reason is that *George* is a closer antecedent than *the regent* and hence the referential properties of reflexive pronouns are subject to the Locality Principle: a reflexive pronoun can only refer to an antecedent within a limited area of the sentence known as a local domain.

The Locality Principle is clearly not just a rule which tells us how to form a particular construction in English; it is something far more general, which applies to phenomena such as reference as well as to movements. It doesn't lead to the formation of any specific construction, but applies to many constructions. The key difference between a rule and a principle is that, while a rule is construction specific, a principle applies to constructions across the board. Chomsky's claim is that human grammars are constructed entirely of principles, not of rules. Specific constructions and the rules for them, such as those we have looked at, are the result of the complex interaction of a number of general principles, including the Locality Principle.

Moreover, principles are universal and so applicable in all human languages. The Locality Principle can indeed limit movement and reference in all languages (with some degree of parameterization – as seen in the next section). For example inversion phenomena, whereby a verbal element moves to the front of a clause, can be found in numerous languages, including German and French. Constructions involving inversion in these languages conform to the Locality Principle:

(33) a. Liest Hans das Buch?
(reads Hans the book)
Does Hans read the book?
b. Hat Hans das Buch gelesen?
(has Hans the book read)
Has Hans read the book?
c. * Gelesen Hans das Buch hat?

In these German examples, the verb moves to the front of the sentence to form a yes-no question (33a) and when there is an auxiliary and a main verb, the auxiliary undergoes the movement (33b). However, in this case, the main verb cannot move as the movement of the auxiliary is the shorter. Hence, verb movement in German conforms to the Locality Principle. The same is true of French:

(34) a. Quand lit-il le livre?
(when reads-he the book)
When does he read the book?
b. Quand a-t-il lu le livre?
(when has-he read the book)
c. *Quand lu-t-il a le livre?

As in English, when a wh-element moves to the front of the clause to form a question, we get inversion of the verb as well. The main verb can invert with the

subject, as in (34a), when there is no auxiliary. When there is an auxiliary, it is this that inverts (34b) and the main verb cannot (34c). The movement of the auxiliary is the shorter one and hence verb movement in French also conforms to the Locality Principle.

Subject raising also occurs in other languages. But it is always a short movement, in accordance with the Locality Principle.

Wh-movement is similarly restricted in other languages, demonstrating the universality of the Locality Principle. In Hungarian a wh-element moves to the front of the verb, but it cannot move out of a clause that starts with a relative pronoun:

(35) a. János találkozott Péterrel.
(John met Peter-with)
John met Peter.

b. János kivel találkozott – ?
(John wh-with met)
Who did John meet?

(36) a. János találkozott az emberrel [aki látta Pétert].
(John met the man-with who saw Peter)
John met the man who saw Peter.

b. * János kit találkozott az emberrel [aki látta –]?
(John who(m) met the man-with who saw)
* Who did John meet the man who saw?

Clearly in Hungarian short movements are grammatical while long movements are ungrammatical, in accordance with the Locality Principle.

Finally, the effects of the Locality Principle in reference phenomena can also be seen in other languages than English. In the following French example, the reflexive *se* can refer to *Jean* but not to *Pierre*:

(37) Pierre dit que Jean se regarde dans la glace.
(Peter says that John himself looks at in the mirror)
Peter says John looks at himself in the mirror.

Similarly in the following Arabic example, the reflexive can refer to the nearby *Zaydun* but not to the distant *Ahmed*:

(38) Qala Ahmed ?anna zaydun qatala nafsahu.
(said Ahmed that Zaid killed himself)
Ahmed said that Zaid killed himself.

The Locality Principle applies in this case too. The Locality Principle is then a universal principle that applies to a wide variety of constructions in many languages: it is part of UG. Of course, the discussion given so far has been necessarily superficial and locality phenomena are more properly analysed as the result

of other universal principles, as we will see. The point is, however, whatever principles are involved in accounting for the phenomena discussed above, they are not construction specific, like grammatical rules, and they are universal, that is to say, applicable to all languages.

THE LOCALITY PRINCIPLE

Principles: these are general conditions that hold for many different constructions. The claim of P&P Theory is that human languages consist of principles with no construction specific rules.

Locality: this principle is a property of linguistic processes which restricts their application to a limited part of the sentence. This then forces movements in all languages to be local: they must be short:

i) It seems he is likely – to leave.

ii) * He seems it is likely – to leave.

The principle also limits the reference of reflexive pronouns to a local domain universally:

iii) Qala Ahmed ?anna zaydun qatala nafsahu.
(said Ahmed that Zaid killed himself)

2.1.6 *An example of a parameter: the head parameter*

Locality seems common to all languages. Yet languages obviously differ in many ways; if knowledge of language consisted solely of invariant principles, all human languages would be identical. To see how the theory captures variation between languages, let us take the example of the head parameter, which specifies the order of certain elements in a language.

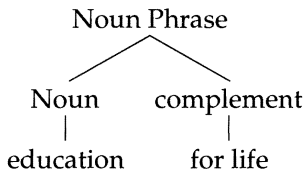
A crucial innovation to the concept of phrase structure that emerged in the early 1970s (Chomsky, 1970) was the claim that all phrases have a central element, known as **a head**, around which other elements of the phrase revolve and which can minimally stand for the whole phrase. Thus the VP *drew an elephant* has a head Verb *drew*; the NP *the child* has a head Noun *child*; a PP such as *by the manager* has a head Preposition *by*; and so on for all phrases. This enabled the structure of all phrases to be seen as having the same properties of head plus other elements, to be expanded in chapter 3. The aim behind this, as always, is to express generalizations about the phrase structure of all human languages rather than features that are idiosyncratic to one part of language or to a single language.

An important aspect of language variation concerns the location of the head in relationship to other elements of the phrase called **complements**. The head of the phrase can occur on the left of a complement or on its right. So in the NP:

(39) education for life

the head Noun *education* appears on the left of the complement *for life*:

(40)

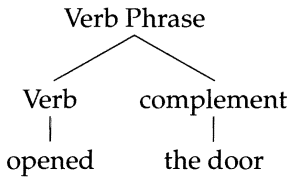


In the VP:

(41) opened the door

the head Verb *opened* appears on the left of the complement *the door*:

(42)

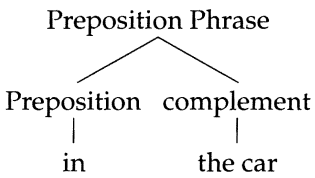


Similarly in the PP:

(43) in the car

the head Preposition *in* appears on the left of the complement *the car*:

(44)

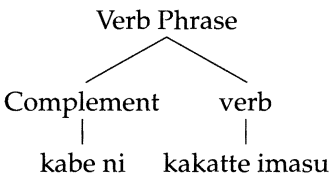


Japanese is very different. In the sentence:

(45) E wa kabe ni kakatte imasu.
(picture wall on is hanging)
The picture is hanging on the wall.

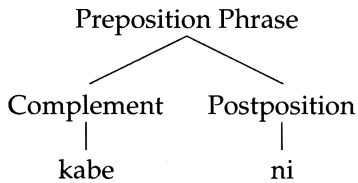
the head Verb *kakatte imasu* occurs on the right of the Verb complement *kabe ni*:

(46)



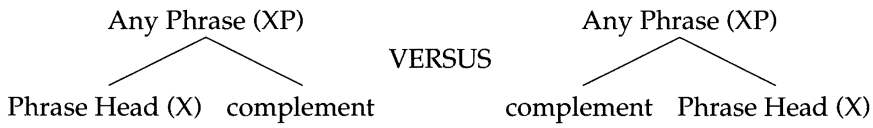
and the *Post*position *ni* (on) comes on the right of the PP complement *kabe*:

(47)



There are thus two possibilities for the structure of phrases in human languages: head-left or head-right:

(48)



Chomsky (1970) suggested that the relative position of heads and complements needs to be specified only once for all the phrases in a given language, creating the 'X-bar syntax' to be described in the next chapter. Rather than a long list of individual rules specifying the position of the head in each phrase type, a single generalization suffices: 'heads are last in the phrase' or 'heads are first in the phrase'. If English has heads first in the phrase, it is unnecessary to specify that verbs come on the left in Verb Phrases, as in:

(49) liked him

or Prepositions on the left in Preposition Phrases, as in:

(50) to the bank

Instead the order of the elements in all English phrases is captured by a single head-first generalization.

Japanese can be treated in the same way: specifying that Japanese is a head-last language means that the Verb is on the right:

(51) Nihonjin desu.
(Japanese am)
(I) am Japanese.

and that it has Postpositions that come after the complements rather than prepositions:

(52) Nihon ni
(Japan in)
in Japan

And the same for other languages. Human beings know that phrases can be either head-first or head-last (alternatively known as head-initial and head-final); an English speaker has learnt that English is head-first; a Japanese speaker that Japanese is head-last, and so on. The word order variation between languages can now be expressed in terms of whether heads occur first or last in the phrase. The variation in the order of elements amounts to a choice between head-first and head-last. UG captures the variation between languages in terms of a limited choice between two or so possibilities – the **head parameter**. The settings for this parameter yield languages as different as English and Japanese. 'Ideally we hope to find that complexes of properties differentiating otherwise similar languages are reducible to a single parameter, fixed in one or another way' (Chomsky, 1981a, p. 6).

The argument here first showed that the speaker of a language knows a single fact that applies to different parts of the syntax, e.g. the phrases of the language consistently have heads to the left. Then it postulated a parameter that all languages have heads either to the left or to the right of their complements. Unlike the universal necessity for locality in movement, the head parameter admits a limited range of alternatives: 'head-first' or 'head-last', depending on the particular language.

Thus alongside the unvarying principles that apply to all languages, UG incorporates 'parameters' of variation; a language 'sets' or 'fixes' the parameters according to the limited choice available. English sets the head parameter in a particular way so that heads of phrases come on the left; Japanese sets the parameter so that they come on the right. This account of the head parameter inevitably simplifies a complex issue; alternative approaches to the word order within the phrase are discussed in more detail in chapter 4. In particular it should be noted both that there are some exceptions to the notion that all phrases have the same head direction in a particular language (for example Hungarian has postpositions but head-initial NPs), and that there are claims that head direction may link to the general performance requirement for language processing that 'lighter' elements precede 'heavier' (Hawkins, 2003).

THE HEAD PARAMETER

Nature: a parameter of syntax concerning the position of heads within phrases, for example Nouns in NPs, Verbs in VPs etc.

Definition: a particular language consistently has the heads on the same side of the complements in all its phrases, whether head-first or head-last.

Examples:

English is head-first:

in the bank: Preposition head first before the complement NP in a PP

amused the man: Verb head first before the complement NP in a VP

Japanese is head-last:

Watashi wa nihonjin desu (I Japanese am): V (*desu*) head last in a VP

Nihon ni (Japan in): P (*ni*) head last in a PP

EXERCISE 2.3

Knowing the head direction of a language can provide you with important information about its structure. Here is some vocabulary in different languages; make trees for possible phrases and sentences in these languages.

Arabic (head-first): *rajol* (man); *khobz* (bread); *ahabba* (liked)
Chinese (head-final): *zai* (in); *chuan* (ship); *shang* (on); *gou* (dog)
Japanese (head-final): *on'na no ko* (girl); *okaasan* (mother); *shiawasema* (happy); *miru* (see)
Italian (head-first): *squadra* (team); *scudetto* (championship); *vincere* (win)
(Note that to get trees for whole sentences you may need information about where the subject occurs, not given here.)
Does the fact you can write trees for languages you may not know show that the head direction parameter is in fact universal?

2.2 Language acquisition

2.2.1 *The language faculty*

Already lurking in the argument has been the assumption that language knowledge is independent of other aspects of the mind. Chomsky has often debated the necessity for this separation, which he regards as 'an empirical question, though one of a rather vague and unclear sort' (Chomsky, 1981b, p. 33). Support for the independence of language from the rest of the mind comes from the unique nature of language knowledge. The Locality Principle does not necessarily apply to all aspects of human thinking; it is not clear how UG principles could operate in areas of the mind other than language. People can entertain mathematical or logical possibilities that are not 'local'; they can even imagine linguistic constructions that do not conform to locality by means of their logical faculties, as the asterisked sentences of linguists bear witness. Nor do the principles of UG seem to be a prerequisite for using language as communication; it might be as easy to communicate by means of questions that reverse the linear order of items as by questions that are based on allowable movement; a language without restrictions on the positions of heads in phrases might be easier to use.

Further arguments for independence come from language acquisition; principles such as Locality do not appear to be learnable by the same means that, say, children learn to roller-skate or to do arithmetic; language acquisition uses special forms of learning rather than those common to other areas.

Chomsky does not, however, claim that the proposal to integrate language with other faculties is inconceivable, simply that the proposals to date have been inadequate; 'since only the vaguest of suggestions have been offered, it is impossible, at present, to evaluate these proposals' (Chomsky, 1971, p. 26). In the absence of more definite evidence, the uniqueness of language principles such as Locality points to an autonomous area of the mind devoted to language

knowledge, a 'language faculty', separate from other mental faculties such as mathematics, vision, logic, and so on. Language knowledge is separate from other forms of representation in the mind; it is not the same as knowing mathematical concepts, for example.

Thus the theory divides the mind into separate compartments, separate modules, each responsible for some aspect of mental life; UG is a theory only of the language module, which has its own set of principles distinct from other modules and does not inter-relate with them. This contrasts with cognitive theories that assume the mind is a single unitary system, for example 'language is a form of cognition; it is cognition packaged for purposes of interpersonal communication' (Tomasello, 1999, p. 150). The separation from other faculties is also reflected in its attitude to language acquisition; it does not see language acquisition as dependent on either 'general' learning or specific conceptual development but sui generis. Thus it conflicts with those theories that see language development as dependent upon general cognitive growth; Piaget for instance argues for a continuity in which advances in language development arise from earlier acquired cognitive processes (Piaget, 1980).

In some ways the theory's insistence on modularity resembles the nineteenth-century tradition of 'faculty' psychology, which also divided the mind into autonomous areas (Fodor, 1983). The resemblance is increased by a further step in the argument. We speak of the body in terms of organs – the heart, the lungs, the liver, etc. Why not talk about the mind in terms of mental organs – the logic organ, the mathematics organ, the common-sense organ, the language organ? 'We may usefully think of the language faculty, the number faculty, and others as "mental organs", analogous to the heart or the visual system or the system of motor coordination and planning' (Chomsky, 1980a, p. 39). The mistake that faculty psychology made may have been its premature location of these organs in definite physical sites, or 'bumps', rather than its postulation of their existence. On the one hand: 'The theory of language is simply that part of human psychology that is concerned with one particular "mental organ", human language' (Chomsky, 1976, p. 36); on the other: 'The study of language falls naturally within human biology' (Chomsky, 1976, p. 123). For this reason the theory is sometimes known as the biological theory of language (Lightfoot, 1982); the language organ is physically present among other mental organs and should be described in biological as well as psychological terms, even if its precise physical location and form are as yet unknown. 'The statements of a grammar are statements of the theory of mind about the I-language, hence statements about the structures of the brain formulated at a certain level of abstraction from mechanisms' (Chomsky, 1986a, p. 23). The principles of UG should be relatable to physical aspects of the brain; the brain sciences need to search for physical counterparts for the mental abstractions of UG – 'the abstract study of states of the language faculty should formulate properties to be explained by the theory of the brain' (Chomsky, 1986a, p. 39), i.e. question (iv) on page 12 above; if there are competing accounts of the nature of UG, a decision between them may be made on the basis of which fits best with the structure of brain mechanisms.

The language faculty is concerned with an attribute that all people possess. All human beings have hearts; all human beings have noses. The heart may be

damaged in an accident, the nose may be affected by disease; similarly a brain injury may prevent someone from speaking, or a psychological condition may cause someone to lose some aspect of language knowledge. But in all these cases a normal human being has these properties by definition. 'This language organ, or "faculty of language" as we may call it, is a common human possession, varying little across the species as far as we know, apart from very serious pathology' (Chomsky, 2002, p. 47). Ultimately the linguist is not interested in a knowledge of French or Arabic or English but in the language faculty of the human species. It is irrelevant that some noses are big, some small, some Roman, some hooked, some freckled, some pink, some spotty; the essential fact is that normal human beings have noses that are functionally similar. All the minds of human beings include the principle that movement is local; it is part of the common UG. It is not relevant to UG Theory that English employs locality in one way, French in another, German in another: what matters is what they have in common.

The words 'human' or 'human being' have frequently figured in the discussion so far. The language faculty is indeed specific to the human species; no creature apart from human beings possesses a language organ. The evidence for this consists partly of the obvious truth that no species of animal has spontaneously come to use anything like human language; whatever apes do in captivity, they appear not to use anything like language in the wild (Wallman, 1992). Some controversial studies in recent years have claimed that apes in particular are capable of being taught languages. Without anticipating later chapters, it might be questioned whether the languages used in these experiments are fully human-like in incorporating principles such as Locality; they may be communication systems without the distinctive features of human language. It may be possible to learn them via other faculties than language at the animal's disposal; in a human being some aspects of language may be learnable by some other means than the language faculty; a linguist may know that movement in German is local without having any knowledge of German simply because movement in all languages is local. Similarly patterns of learning used by the animal for other purposes may be adapted to learning certain aspects of language. The danger is that this argument could evade the issue: how is it possible to tell 'proper' language knowledge gained via the language faculty from 'improper' language knowledge gained in some other way, an issue that is particularly important to second language acquisition? (This will become important in chapter 6.) Presumably only by returning to the first argument: if it embodies principles of UG and has been acquired from 'natural' evidence, then it is proper. None of the systems learnt by animals seems proper in this sense, either because they fail to reflect abstract features of language or because they are artificially 'taught'. Or, indeed, because they fail to be creative in the sense seen in the last chapter: 'animal communication systems lack the rich, expressive and open-ended power of human language' (Hauser et al., 2002, p. 1570).

The species-specificity of UG nevertheless raises difficult questions about how it could have arisen during evolution; Piaget, for instance, claims 'this mutation particular to the human species would be biologically inexplicable' (Piaget, 1980, p. 31). However, 'the fact that we do not know how to give serious evolutionary explanation of this is not surprising; that is not often possible beyond simple cases' (Chomsky, 2000a, p. 50). While the possession of language itself clearly confers

an immense advantage on its users over other species, why should Locality confer any biological advantage on its possessor? Indeed one puzzle is why there are different human languages: it would seem advantageous if the whole species spoke the same language. Presumably our lack of distance from human languages makes them appear so different to us; the differences between Japanese and English might seem trivial to a non-human alien. 'The Martian scientist might reasonably conclude that there is a single human language, with differences only at the margins' (Chomsky, 2000b, 7).

In the 2000s Chomsky has been developing a slightly different way of thinking of the language faculty in association with evolutionary biologists (Hauser et al., 2002). He now makes a distinction between the broad faculty of language (FLB) and the narrow faculty of language (FLN). The FLN 'is the abstract linguistic computational system alone, independent of the other systems with which it interacts and interfaces' while the FLB includes as well 'at least two other organism-internal systems, which we call "sensory-motor" and "conceptual-intentional"' (Hauser et al., 2002, pp. 1570–1). This proposal seems to restrict the language faculty severely, leaving little that is unique to language, possibly only the core property of recursion, which allows rules to call upon themselves and will be discussed in chapter 3.

The distinction between FLB and FLN has proved highly controversial in that it seems to concede much of the ground to Chomsky's opponents and leave very little that is peculiar to language. The narrow language faculty is now a small unique area; the broad language faculty that includes all the language-related systems is no longer unique and much of it may be shared with the animal kingdom. Indeed, if recursion is all that is left, some have pointed out that this too is shared with other faculties – even Hauser et al. (2002) mention it is a property of natural numbers – and that not all human languages have it (Everett, 2005). On the other hand the FLN seems to have thrown the baby out with the bathwater if it casts vocabulary and phonology out of the core language faculty. For a taste of the raging debate over this issue, readers are referred to the critique by Pinker and Jackendoff (2005) and the answer by Fitch et al. (2005).

EXERCISE 2.4

1 Compare from your experience the faculties of language and mathematics – what knowledge you have in them, how you learnt them, and how you use them in everyday life. Does this convince you they are separate or overlap?

2 Here are some of the 27 faculties proposed by the phrenologist Franz Gall in the 1790s. Which might be seen as faculties today? Which might interface with the faculty of language?

impulse to propagation	tenderness for the offspring,
murder, carnivorousness	sense of cunning
faculty of language	sense for sounds, musical talent
arithmetic, counting, time	metaphysical perspicuity
poetic talent	recollection of persons
mimic	perseverance, firmness

THE LANGUAGE FACULTY

- is where the knowledge of language is stored in the individual mind
- is common to all human beings
- is independent of other faculties such as mathematics
- has unique properties of its own like Locality or recursion not shared with other faculties
- is unique to the human species, at least in the narrow sense
- can be thought of as a 'mental organ' that 'grows'

2.2.2 *States of the language faculty*

Let us now relate the language faculty to the acquisition of language, always central to the UG Theory. The language faculty can be thought of as a state of the mind, containing whatever the speaker knows at a particular point in time, the sum of all their knowledge of language, variously called a grammar or an I-language: 'The internal language, in the technical sense, is a state of the faculty of language' (Chomsky, 2002, p. 48). The language faculty then comprises a computational system with principles, parameters and a lexicon all fleshed out for a particular language, as seen on the right of figure 2.1.

Figure 2.1 The states model of the development of the language faculty

But children are not born with the knowledge of all the lexical items in the language. In the initial state, the parameters have not been set, the lexical items have not been learnt etc., as seen in the left of figure 2.1; this is the language faculty with the minimal contents – whatever aspects of language are intrinsic to the human mind, that is to say UG. The two extreme states of the language faculty are then the final state when the mind knows a complete I-language and the initial state when it knows only the principles. Language acquisition comes down to how the human language faculty changes from the initial to the final state, how children acquire all the knowledge of language seen in the adult.

Let us sum up this section in Chomsky's own words: 'the language organ is the *faculty of language* (FL); the theory of the initial state of FL, an expression of the genes, is *universal grammar* (UG); theories of states attained are *particular grammars*; the states themselves are *internal languages*, "languages" for short' (Chomsky, 2005b, p. 145). The child's mind starts in the initial state of the language faculty (alias UG); this contains only whatever is genetically determined about language – principles such as Locality etc. This in itself takes a controversial position about the innate elements in language, as we shall see. The language faculty achieves adult knowledge of language, complete with parameter settings and lexicon for

a particular language, by getting certain types of information about the structures and vocabulary of the language it is exposed to. Put another way, UG gets instantiated as the knowledge of a particular language; the language faculty still has the same structure but the syntactic peculiarities and lexical items of a particular language have become attached to it. In between the beginning and end positions the language faculty evolves through a number of states, each of them a possible language of its own. So language acquisition amounts to the child's mind fleshing out the skeleton of language knowledge already present in its mind with the material provided by the environment. This 'states' view of acquisition sees the whole language faculty as involved: the language faculty incorporates the information about the specific language within itself to get a grammar of a particular language.

We can now elaborate this in figure 2.2. The grammar is a state of UG, the faculty of language in the human mind, not a product of UG.

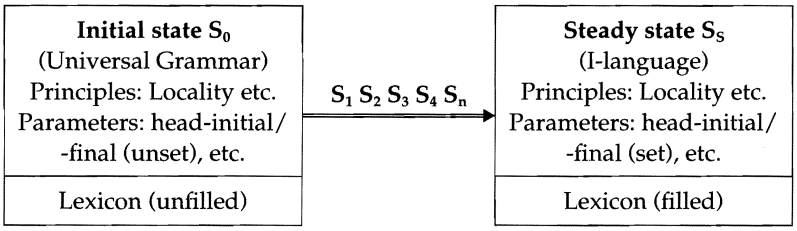


Figure 2.2 The development of the language faculty: zero to final states

In the beginning of language acquisition is the mind of the new-born baby who knows no language, termed the **initial (zero) state** or $\mathbf{S_0}$, containing nothing but UG itself. At the end is the mind of the adult native speaker with full knowledge of the language, including the principles, parameter settings and lexicon. This final state is, to all intents and purposes, static; the speaker may become more or less efficient at using language or may add or lose a few vocabulary items – *step-change* suddenly entered British people's vocabulary after a single politician's remark – but linguistic competence is essentially complete and unchanging once it has been attained. To sum up in Chomsky's words, the faculty of language 'has a genetically-determined initial state S_0, which determines the possible states it can assume' (Chomsky, 2001b, p. 1).

While for many purposes it is convenient to look at just the initial and final states of the acquisition process, the language faculty goes through many intervening states while children are acquiring the language. 'The initial state changes under the triggering and shaping effect of experience, and internally determined processes of maturation, yielding later states that seem to stabilise at several stages, finally at about puberty' (Chomsky, 2002, p. 85). Figure 2.2 therefore shows a series of states – $S_1, S_2, S_3, S_4 \ldots S_n$ – intervening between the initial state S_0 and the steady state S_S, each of them a possible state of the language faculty incorporating a knowledge of principles, parameters and the lexicon. Acquiring language means progressing from not having any language, S_0, to having full competence, S_s.

2.2.3 Behaviourism

To some extent, Chomsky's ideas are now so taken for granted that their originality has been obscured. For, prior to Chomsky's work of the late fifties, language was not considered to be what people knew but how they behaved, as incorporated in the structuralist linguistics tradition best known from Bloomfield's book *Language* (Bloomfield, 1933). Bloomfield saw language acquisition as initiated by the child more or less accidentally producing sounds such as *da*; these sounds become associated with a particular object such as a doll because of the parents' reactions, so that the child says *da* whenever a doll appears; then the child learns to talk about *doll* even when one is not present – 'displaced speech'. The adult is a crucial part of the process; the child would never learn to use *da* for 'doll' without the adult's reaction and reinforcement. This Bloomfieldian version of language acquisition was the commonplace of linguistics before Chomsky.

What Chomsky specifically repudiated, however, was the more sophisticated behaviourist theory of B. F. Skinner, put forward in *Verbal Behavior* (Skinner, 1957), a sympathetic account of which can be found in Paivio and Begg (1981). Skinner rejected explanations for language that were inside the organism in favour of explanations in terms of outside conditions. Language is determined by *stimuli* consisting of specific attributes of the situation, by *responses* the stimuli call up in the organism, and by *reinforcing stimuli* that are their consequences. Thus the object 'doll' acts as a stimulus for the child to respond *Doll*, which is reinforced by the parents saying *Clever girl!* or by them handing the child the doll. Or the child feels thirsty – the stimulus of 'milk-deprivation' – responds by saying *Milk*, and is reinforced with a glass of milk. As with Bloomfield, language originates from a physical need and is a means to a physical end. The parents' provision of reinforcement is a vital part of the process.

Chomsky's classic critique of Skinner in his review of *Verbal Behavior* (Chomsky, 1959) presaged many of his later ideas. Chapter 1 here introduced the key Chomskyan notion of *creativity*; people regularly understand and produce sentences that they have never heard before, say *Daeman faxed into solidity near Ada's home* (Simmons, 2004). How could they be acting under the control of stimuli, when, outside science fiction, they have never encountered the concept of a human being being faxed or even seen *faxed* used as an intransitive verb? To take Chomsky's examples, you do not need to have experienced the situation before to take appropriate action if someone says *The volcano is erupting*.

Nor is a stimulus usually as simple and unambiguous as milk-deprivation or a volcano erupting. Chomsky imagines the response of a person looking at a painting. They might say *Dutch* or *Clashes with the wallpaper, I thought you liked abstract art, Never saw it before, Tilted, Hanging too low, Beautiful, Hideous, Remember our camping trip last summer?*, or anything else that comes to mind. One stimulus apparently could have many responses. There can be no certain prediction from stimulus to response. One of the authors went to a supermarket in Kassel, prepared with language teaching clichés for conversations about buying and selling; the only German that was addressed to him was:

(53) Könnten Sie mir bitte den Zettel vorlesen, weil ich meine Brille zu Hause vergessen habe?
Could you please read the label to me because I've left my glasses at home?

In other words, human language is unpredictable from the stimulus. The important thing about language is that it is *stimulus-free*, not *stimulus-bound* – we can say anything anywhere without being controlled by precise stimuli.

It is also hard to define what reinforcement means in the circumstances of children's actual lives, rather than in the controlled environment of a laboratory. The child rarely encounters appropriate external rewards or punishment; 'it is simply not true that children can learn language only through "meticulous care" on the part of adults who shape their verbal repertoire through careful differential reinforcement' (Chomsky, 1959, p. 42). The burden of Chomsky's argument is, on the one hand, that Skinnerian theory cannot account for straightforward facts about language, on the other, that the apparently scientific nature of terms such as 'stimulus' and 'response' disguises their vagueness and circularity. How do we know the speaker was impressed by the 'Dutchness' of the painting rather than by some other quality? Only because they said *Dutch*: We are discovering the existence of a stimulus from the response rather than predicting a response from a stimulus. This early demolition of Skinner still remains Chomsky's main influence on psychology, rather than his later work; introductions to psychology seldom mention post-1965 writing.

EXERCISE 2.5 People have often distinguished language-like behaviour from language behaviour. If you took the following activities, which could you say were true language, which language-like?

- riding a bicycle
- doing mental arithmetic
- reading a map
- improvising an epic poem
- praying
- miming how to do a task
- exclaiming 'ouch' when you hit your finger
- finding your way in a maze

Do you feel this shows a clear distinction between language and non-language processing by the individual, as Chomsky claims, or that the same processes are involved in language as in other cognitive operations?

2.2.4 *The Language Acquisition Device and levels of adequacy*

A more familiar way of thinking about language acquisition put forward by Chomsky is in terms of a **Language Acquisition Device (LAD)**. Chomsky (1959, p. 58) provided the germ out of which this model grew; 'in principle it may be

possible to study the problem of determining what the built-in structure of an information-processing (hypothesis-forming) system must be to enable it to arrive at the grammar of a language from the available data in the available time'. Chomsky (1964) put this metaphorically as a black box problem. Something goes into a black box, something comes out. By looking at the input and the output, it is possible to arrive at some understanding of the process concealed inside the box itself. Suppose we see barley and empty bottles going in one door of a Speyside distillery, crates of Scotch whisky coming out the other; we can deduce what is going on inside by working out what must be done to the barley to get whisky. Given a detailed analysis of the whisky and of the barley, we could deduce the processes through which one is transformed into the other.

Children hear a number of sentences said by their parents and other caretakers – the **primary linguistic data**; they process these within their black box, the LAD, and they acquire linguistic competence in the language – a 'generative grammar' in their minds. We can deduce what is going on inside the child's LAD by careful examination and comparison of the language input that goes in – the material out of which language knowledge is constructed – and the knowledge of language that comes out – the generative grammar. 'Having some knowledge of the characteristics of the acquired grammars and the limitations on the available data, we can formulate quite reasonable and fairly strong empirical hypotheses regarding the internal structure of the language acquisition device that constructs the postulated grammars from the given data' (Chomsky, 1972a, p. 113). In the case of the whisky, we could go into the distillery to check on our reasoning and see just what is going on inside; it is not of course possible to open the child's mind to confirm our deductions in the same fashion; the black box of the mind cannot be opened.

The model of this process proposed by Chomsky (1964) was as shown in figure 2.3, adapted slightly.

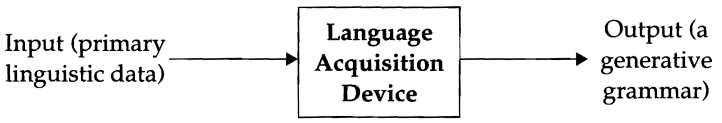


Figure 2.3 The Language Acquisition Device model of first language acquisition

The LAD is 'a procedure that operates on experience acquired in an ideal community and constructs from it, in a determinate way, a state of the language faculty' (Chomsky, 1990, p. 69). Since essentially all human children learn a human language, it has to be capable of operating for any child anywhere; it must tackle the acquisition of Chinese as readily as the acquisition of English, the acquisition of Russian as readily as that of Sesotho.

The LAD conceptualization was a powerful metaphor for language acquisition within UG theory. McCawley (1992) indeed insists that it is true of *any* theory of language acquisition, not just UG. It embodied the central tenet of the theory by treating language acquisition as the acquisition of knowledge. While this may now seem obvious, it was nevertheless at odds with the accounts of language

acquisition based on behaviour that had been provided before the 1960s. The LAD metaphor said that it was not how children behaved that mattered; it was not even what they actually said: it was what they *knew*.

The LAD led to a neat way of putting the goals of linguistics in terms of three 'levels of adequacy' (Chomsky, 1964), foreshadowing the goals of linguistics described in chapter 1.

- **Observational adequacy** is the first level that a linguistic theory has to meet: a theory is observationally adequate if it can predict grammaticality in samples of language, that is to say in the primary linguistic data of adult speech as heard by the child, otherwise called the input to the LAD.
- **Descriptive adequacy** is the second level: a theory achieves descriptive adequacy if it deals properly with the linguistic competence of the native speaker, i.e. the generative grammar output from LAD.
- **Explanatory adequacy** is the third level: a theory is explanatorily adequate if the linguistics theory can provide a principled reason why linguistic competence takes the form that it does, i.e. if it can explain the links between linguistic competence and primary linguistic data that are concealed within the LAD itself.

Explanatory adequacy was presented as a method of deciding between two descriptions of linguistic competence both of which seem descriptively and observationally adequate – mathematically speaking a not unlikely event given that an infinite number of descriptively adequate grammars is possible. The preferred description of the output grammar, whenever there is a choice, is the one that children can learn most easily from the language data available to them. A proper linguistic theory has then to meet ' "the condition of explanatory adequacy": the problem of accounting for language acquisition' (Chomsky, 2000a, p. 13).

The 1964 LAD model can be accommodated within P&P Theory to some extent. The LAD itself is synonymous with the language faculty, i.e. UG. 'In general, the "language acquisition device" is whatever mediates between the initial state of the language faculty and the states it can attain, which is another way of saying it is a description of the initial state' (Chomsky, 2000a, p. 55). The linguistic competence output that emerges from LAD consists of a grammar couched in principles and parameters form; the knowledge that the child needs to acquire consists of the setting of the head parameter, etc. The grammar contains the appropriate parameter settings and has thousands of lexical entries specifying how each word can behave in the sentence. The translation into UG terms can be expressed in figure 2.4, now conventional in slight variations in the literature, for example Haegeman (1994, p. 15) and Atkinson (1992, p. 43).

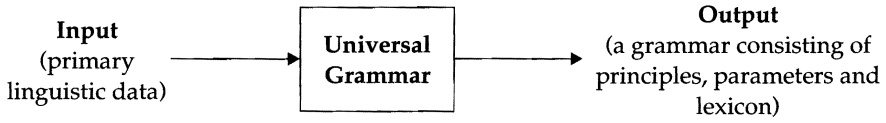


Figure 2.4 The Universal Grammar model of first language acquisition

The UG revision to the LAD model alters the relative importance of the levels of adequacy. Explanatory adequacy had seemed an ideal but fairly distant goal; most energy had gone into the descriptions themselves. Acquisition seldom had a real role in deciding on the right linguistic theory. The UG version with principles and parameters, however, integrated acquisition with the description of grammar by making explanatory adequacy central; the description of the grammar goes hand in hand with the explanation of how it is learnt. In principle any element in the grammar has to be justified in terms of acquisition; any principle or parameter that is proposed for the speaker's knowledge of syntax has to fit into an account of acquisition. So all the technical apparatus of P&P Theory must ultimately be integrated with the theory of language acquisition.

Chomsky and other researchers essentially switch between these two metaphors of the changing state and the input/output box, assuming them to be equivalent. However, the changing state metaphor sees UG itself as the initial grammar, which evolves over time; the input/output metaphor implies an unchanging UG producing a series of grammars distinct from itself. As we see below, the discussion of second language acquisition reveals some contradictions between the two metaphors (Cook, 1993).

LEVELS OF ADEQUACY

- **Observational adequacy**: faithfulness to the primary linguistic data of adult speech
- **Descriptive adequacy**: faithfulness to the linguistic competence of the native speaker
- **Explanatory adequacy**: faithfulness to the acquisition of linguistic competence

2.2.5 *The poverty-of-the-stimulus argument*

The black box LAD model led to further interesting ideas about acquisition. To return to the distillery, barley is going in and whisky is coming out, but where does the water come from that makes up 43 per cent of distillery-strength Scotch? Anything that comes out of the distillery that has not been seen to go in must have originated within the distillery itself. So there is presumably a source of water inside the distillery that the observer can't see from the outside.

Suppose, however, something comes out of the LAD that didn't go in: where could this ingredient come from? This is the conundrum called 'Plato's problem', which is at the heart of Chomskyan ideas of language acquisition: 'How do we come to have such rich and specific knowledge, or such intricate systems of belief and understanding, when the evidence available to us is so meagre?' (Chomsky, 1987). The answer is that much of our linguistic knowledge must come from the internal structure of the mind itself, in terms of the distillery metaphor a hidden well. If the adult's grammar S_s incorporates principles that could not be constructed

from the primary linguistic data then they must have been added by the mind itself. The things that are missing from the input are added by the mind: the black box is not just processing the input but contributing things of its own.

Our knowledge of language is complex and abstract; the experience of language we receive is limited. Human minds could not create such complex knowledge on the basis of such sparse information. It must therefore come from somewhere other than the limited evidence they encounter. Plato's own solution was to say that the knowledge originated from memories of prior existence; Chomsky's solution is to invoke innate properties of the mind: it doesn't have to be learnt because it is already there. This argument has a clear and simple form: on the one hand there is the complexity of language knowledge, on the other there are the impoverished data available to the learner; if the child's mind could not create language knowledge from the data in the surrounding environment, given plausible conditions on the type of language evidence available, the source must be within the mind itself. This is therefore known as the **poverty-of-the-stimulus argument**, meaning that the data in the stimulus are too meagre to justify the knowledge that the mind builds out of them.

Let us go over the poverty-of-the-stimulus argument informally before putting it in more precise terms. Part of the linguistic competence of a native speaker is demonstrably the principle of Locality, sketched above. Research reported in Cook (2003) asked for grammaticality judgements on sentences involving different parameters from first language (L1) speakers of English and L2 speakers from several backgrounds. Native speakers of English indeed rejected sentences involving rules that violated Locality in questions, such as:

(54) *Is Sam is the cat that black?

99.6 per cent of the time. But how could they have learnt this judgement from their parents? What clues might children hear that would tell them that English obeys Locality? Children never hear examples of English sentences that violate this principle since these do not exist outside the pages of linguistics books. Nor is it likely that parents correct them when they get such things wrong, partly because children do not produce such errors, partly because parents would probably not know what they were if they did.

Perhaps it is just the sheer unfamiliarity of the sentence that offends them. Yet native speakers encounter new and strange sentences all the time, which they immediately accept as English, even if they do not fully understand them, say:

(55) On later engines, fully floating gudgeon pins are fitted, and these are retained in the pistons by circlips at each end of the pin. (Haynes, 1971, p. 29)

It is not that a sentence that breaches grammatical principles is novel or necessarily incomprehensible: we know it is *wrong*. The child has been provided with no clues that the Locality Principle exists – but nevertheless knows it. The source of the Locality Principle is not outside the child's mind since the environment does not provide any clues to its existence. As this constraint is part of the grammar

that came out of the black box but it did not go in as part of the input, it must be part of the black box itself: the Locality Principle must already have been present in the child's mind. Thus it is an innate aspect of the human language faculty.

There are four steps to the poverty-of-the-stimulus argument (Cook, 1991):

Step A: A native speaker of a particular language knows a particular aspect of syntax. The starting point is then the knowledge of the native speaker. The researcher has to select a particular aspect of language knowledge that the native speaker knows, say the Locality Principle, or any other aspect of language knowledge.

Step B: This aspect of syntax could not have been acquired from the language input typically available to children. The next step is to show that this aspect of syntax could not have been acquired from the primary linguistic data, the speech that the child hears. This involves considering possible sources of evidence in the language the child encounters and in the processes of interaction with parents. These will be itemized in greater detail below.

Step C: We conclude that this aspect of syntax is not learnt from outside. If all the types of evidence considered in Step B can be eliminated, the logical inference is that the source of this knowledge does not lie outside the child's mind.

Step D: We deduce that this aspect of syntax is built in to the mind. Hence the conclusion is that the aspect of syntax must originate within the child's mind. Logically, as the aspect did not enter from without, it must come from within, i.e. be a built-in part of the human language faculty.

Steps C and D are kept distinct here because there could be explanations for things which are known but not learnt other than the innate structure of the mind. Plato's memories of previous existence are one candidate, telepathy and morphogenesis modern ones.

Step A: A native speaker of a particular language knows a particular aspect of syntax.
Step B: This aspect of syntax could not have been acquired from the language input typically available to children.
Step C: We conclude that this aspect of syntax is not learnt from outside.
Step D: We deduce that this aspect of syntax is built in to the mind.

Figure 2.5 Steps in the poverty-of-the-stimulus argument for first language acquisition

The steps of this argument can be repeated over and over for other areas of syntax. Whatever the technical details of the syntax, the argument still holds. It may be, as Rizzi (1990) has argued, that Locality derives from other general principles of syntax. But the point is still true: if no means can be found through which the child can acquire language knowledge from the usual evidence he or she may receive then it must be built in to the mind, however controversial or uncertain the linguist's analysis itself may be. The poverty-of-the-stimulus argument

is fundamentally simple; whenever you find something that the adult knows which the child cannot in principle acquire, it must already be present within the child. Indeed the form of the poverty-of-the-stimulus argument has been used in areas other than language. Several religions for example claim that the world is so beautiful or so complex that it could not have come into existence spontaneously and must therefore be due to a creator; this 'argument by design' was used by Paley (1802, quoted in Gould, 1993) as a stick with which to beat evolutionary theories, revived to some extent by the recent arguments for intelligent design.

The crucial steps in the argument are: first that some aspect of language is indeed part of the native speaker's linguistic competence; second that the child does not get appropriate evidence. In a critical review, Pullum and Scholz (2002, p. 9) have teased out a series of separate claims within the poverty-of-the-stimulus argument, put in their terms as:

- Step A addresses the 'acquirendum' – establishing what the speaker knows;
- Step B concerns the 'lacunae' – what sentences are missing from the language input;
- Step C involves 'inaccessibility' – evidence that the lacuna sentences are not actually available to the learner – and 'positivity' – the unavailability of indirect negative evidence;
- Step D relies on 'indispensability' – the argument that the acquirendum could not be learnt without access to the lacunae. In other words 'if you know X, and X is underdetermined by learning experience, then the knowledge of X must be innate'. (Legate and Yang, 2002, p. 153)

Hence the apparent simplicity and clarity of Chomsky's poverty-of-the-stimulus argument may in fact depend on a number of sub-issues that need to be discussed separately.

EXERCISE 2.6 Try to devise poverty-of-the-stimulus arguments for other areas of human development such as:

- knowledge of gravity
- knowledge of mathematics
- knowledge of cookery
- knowledge of vision

Does this convince you that the argument works for language?

2.2.6 *The Principles and Parameters Theory and language acquisition*

The overall model of language acquisition proposed by Chomsky can be put quite simply. UG is present in the child's mind as a system of principles and parameters. In response to evidence from the environment, the child creates a core grammar S_s that assigns values to all the parameters, yielding one of the allowable

human languages – French, Arabic, or whatever. To start with, the child's mind is open to any human language; it ends by acquiring one particular language. What the language learner must do 'is figure out on the basis of his experience, what options his language has taken at every choice point' (Rizzi, 2004, p. 330). The principles of UG are principles of the initial state, S_0. No language breaches them; since they are underdetermined by what the child hears, they must be present from the beginning. They are not learnt so much as 'applied'; they are automatically incorporated into the child's grammatical competence. The resemblances between human languages reflect their common basis in principles of the mind; Japanese incorporates Locality, as do English and Arabic, because no other option is open to the child. While the discussion here concentrates on the acquisition of syntax, Chomsky extends the argument to include 'fixed principles governing possible sound systems' (Chomsky, 1988, p. 26) and 'a rich and invariant conceptual system, which is prior to any experience' (Chomsky, 1988, p. 32); these guide the child's acquisition of phonology and vocabulary respectively.

Acquiring a language means setting all the parameters of UG appropriately. As we have seen, these are limited in number but powerful in their effects. To acquire English rather than Japanese, the child must set the values for the head parameter, and a handful of other parameters. The child does not acquire rules but settings for parameters, which, interacting with the network of principles, create a core grammar. 'The internalised I-language is simply a set of parameter settings; in effect, answers to questions on a finite questionnaire' (Chomsky, 1991b, p. 41). Rather than a black box with mysterious contents, Chomsky is now proposing a carefully specified system of properties, each open to challenge.

In addition to the core grammar, the child acquires a massive set of vocabulary items, each with its own pronunciation, meaning and syntactic restrictions. While the acquisition of core grammar is a matter of setting a handful of switches, the child has the considerable burden of discovering the characteristics of thousands of words. 'A large part of "language learning" is a matter of determining from presented data the elements of the lexicon and their properties' (Chomsky, 1982, p. 8). So the child needs to learn entries that specify that *sleep* is a Verb which requires a subject; that *give* is a Verb that requires a subject, an object and an indirect object; the referential properties of *himself*; and so on for all the items that make up the mental lexicon of a speaker of English.

As well as those aspects derived from UG principles, the child acquires parts of the language that depart from the core in one way or another, for example the irregular past tense forms in English such as *broke* and *flew*. Grammatical competence is a mixture of universal principles, values for parameters, and lexical information, with an additional component of peripheral knowledge for, say constructions like *the more the merrier*. Some of it has been present in the speaker's mind from the beginning; some of it comes from experiences that have set values for parameters in particular ways and led to the acquisition of lexical knowledge. To sum up in Chomsky's words: 'what we "know innately" are the principles of the various subsystems of S_0 and the manner of their interaction, and the parameters associated with these principles. What we learn are the values of the parameters and the elements of the periphery (along with the lexicon to which similar considerations apply)' (Chomsky, 1986a, p. 150).

Discussion topics

1 What could count as evidence that something is or isn't part of the initial state of the language faculty built in to the human mind?
2 What do *you* think is the proper goal of linguistics?
3 Feyerabend (1975) sees science as based on powerful arguments rather than evidence. Is there a way out of Chomsky's poverty-of-the-stimulus argument?
4 Is any major type of evidence overlooked in the discussion that could account for the child's acquisition of principles of UG?
5 To what extent are the arguments that Chomsky proposed against Skinnerian behaviourism valid against modern psychological approaches to language acquisition such as usage-based acquisition (Tomasello, 2003) or connectionism (Rumelhart and McLelland, 1986)?
6 Recent work has tended to extend the areas of language that animals can cope with, for example showing rats are capable of distinguishing Dutch from Japanese (Toro et al., 2005). Does this affect Chomsky's insistence that language is peculiar to human beings? Or is it simply part of the FLB?

3 Structure in the Government/ Binding Model

The previous chapter introduced three key concepts which have been present in all the stages of development of Chomsky's linguistics since the early 1960s:

- *the lexicon*, which stores all idiosyncratic information about the words of the language, in the form of lexical entries;
- *phrase structure rules*, which combine lexical elements to form basic structures; and
- *movement rules*, which shift elements about in structures (displacement).

The immediate consequence of having movement rules is the recognition that there are two levels at which the structure of any sentence can be described: the level *before* the movement takes place and the level formed *after* the movement has happened. So the sentence:

(1) What film did you see?

clearly requires an underlying level:

(2) You saw what film?

with *what film* in its original position.

At the level after things have been moved about, seen in sentence (1), the elements are sitting in positions which are in closer accord with the linear order in which the sentence is actually pronounced; i.e. after movement *what film* occurs at the beginning of the sentence as it does in speech. For this reason this level was initially called **surface structure**. The level before movement seen in sentence (2) is in some ways more abstract than surface structure, with elements sitting in positions different from those indicated by their usual pronounced order, and as such it represents the analysis of structure deeper in the system. It was therefore originally called **deep structure**. However, the connotations of the terms 'deep' and 'surface' were misleading and the more neutral terms **D-structure** and **S-structure** were later adopted.

The basic form of the grammar, incorporating the relationship of D-structure to S-structure via movement, is represented in figure 3.1,

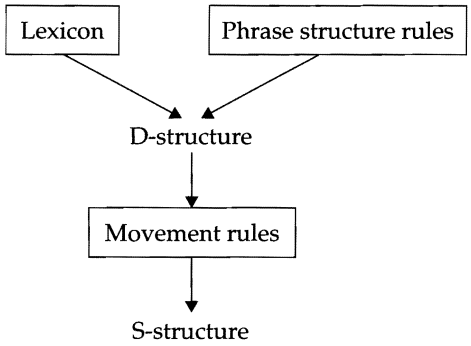


Figure 3.1 The basic form of the grammar 1980s–90s

This then attempts to model the computational system that the mind uses to bridge sounds and meanings, as seen in the models in chapter 2. Its core remained virtually unchanged from the mid-1960s till the 1990s.

3.1 The heart of the Government/Binding Model

This simple model of the grammar brings out ideas that were to have a profound effect on its development in the 1980s. In particular the grammar is split up into parts that have their own specific roles to play, in Government/Binding (GB) Theory known as the **modules**, two of which are shown in figure 3.1. Perhaps confusingly, some of these modules were called theories, e.g. Binding Theory.

This chapter and the next introduce the modules of GB Theory and the phenomena that each was proposed to account for. This chapter considers those modules relevant for D-structure; the next chapter moves on to those relevant to S-structure, i.e. those which interact with movement.

First we need to outline the modular nature of the grammatical system as a whole and the reasons for it being named Principles and Parameters (P&P) Theory, briefly outlined in chapters 1 and 2.

3.2 Modules, principles and parameters

Prior to GB Theory, at the stage of development around 1970 often referred to as Standard Theory, all grammatical rules were thought to belong to one of the three components of the grammar outlined in figure 3.1. Specifically there were phrase structure rules (= structure rules) and transformational rules (= movement rules), familiar in slightly different forms from the early days of *Syntactic Structures* (Chomsky, 1957). However, work throughout the 1970s indicated that other

types of rule were needed. These 'rules' exert a moderating effect on the existing rules, constraining their actions and allowing them to be greatly simplified.

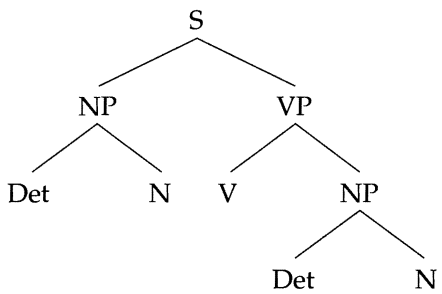
3.2.1 *Phrase structure rules and their restrictions*

Let us start with phrase structure rules. To recap chapter 2, Chomsky's breakthrough in the 1950s was to devise a way of representing grammatical rules as rewrite rules, i.e. instructions about how to generate structures for sentences. A typical basic set of rewrite rules was presented in chapter 2:

(3) (i) S → NP VP
(ii) VP → V NP
(iii) NP → Det N

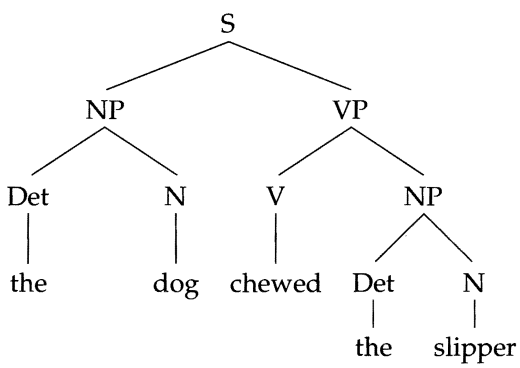
Following these rules, we can generate the structure in (4):

(4)



At the bottom of the tree comes a set of category labels, namely Det, N and V, known as **terminal nodes**, to which we can attach elements from the lexicon of the appropriate category, i.e. an N might be *dog* or *politician*, a V *chew* or *bribe*, a Det *a* or *the*. So this structure can be associated with sentences such as:

(5)



These three rules describing the structure (4) allow for a vast number of sentences, limited only by the number of Nouns, Verbs and Determiners in the lexicon of the language.

We saw in chapter 2 that these rules (3i–iii) are language specific, in as much as they generate the structures of English – other languages need different sets of rewrite rules – and construction specific – each one describes the structure of a particular phrase, such as the Verb Phrase, rather than of all phrases of English – Verb Phrases, Noun Phrases and all the other phrases.

It might be a good idea to collapse some of these rules to produce something more general. For example, alongside the VP rule (3ii) which requires the Verb to have an object, i.e. to be transitive, we also need to deal with intransitive verbs, which lack objects, for example:

(6) The baby sleeps.

The following rule describes sentences where the verb is the only element in the VP:

(7) VP → V

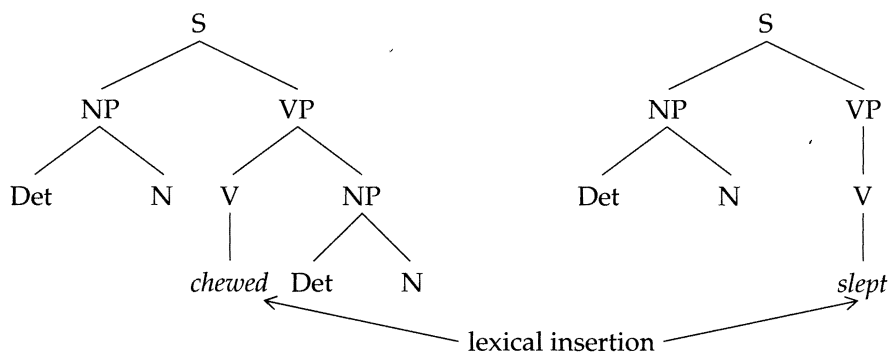
We now have two VP rules (3ii) and (7). However, these two rules can be collapsed into a single, more general, rule by introducing a notation to show that the object in the VP is optional by enclosing it in brackets. The brackets round the NP in (8) therefore indicate its optionality; the object NP may, or may not, be present in the VP:

(8) VP → V (NP)

Two rules (3ii) and (7) have been collapsed into one (8) through the concept of optional elements, yielding two possible outputs of the rule with or without an NP.

Obviously, whether or not the VP has an object depends on which verb is selected to be inserted into the V terminal node: *chew* requires an object, *sleep* does not. This is handled by a lexical insertion rule, which takes note of the structural context in which a particular lexical element can appear. 'Base rules generate D-structures (deep structures) through insertion of lexical items into structures generated by [phrase structure rules], in accordance with their feature structure' (Chomsky, 1981a, p. 5). Thus, if there is an object following the verb in the structure, the lexical insertion rule will only insert transitive verbs such as *chew* into the V terminal node and, if there is no object, the insertion rule will only insert intransitive verbs such as *sleep*:

(9)



The lexical entries for verbs must therefore indicate whether they are intransitive or transitive. This is done by the **subcategorization frame** of the lexical entry in the lexicon, which simply states the contextual conditions under which the lexical item may be inserted into a structure:

(10) *chew*: [__NP]
sleep: [__∅]

The lexical entries in (10) state that *chew* is a transitive verb and hence appears in a position preceding an NP, shown in square brackets with underlined space to indicate the position of the item [__NP]. *Sleep*, however, is an intransitive verb and appears in positions in which there is no following complement (indicated by ∅). We can see that the grammar needs to take into account not only the category of each item – Nouns, Verbs, etc. – but also the subcategories of each category that link it to particular constructions – whether a Verb is followed by an object or not, and so on. Accounting for the behaviour of particular lexical items in sentences then became more and more crucial to the model.

As Chomsky (1970) pointed out, this system contains a number of redundancies and it is also incapable of capturing generalizations across different constructions. For one thing, the subcategorization frames of lexical verbs duplicate the information in the phrase structure rule (8) that VPs can contain verbs and NPs or just verbs: on the one hand is a rule saying a VP contains a Verb and an optional NP; on the other lexical entries that specify *chew* is followed by an NP, *sleep* is not. It is unavoidable that this information be included in the lexicon, as whether or not a particular lexical element subcategorizes for an object is an idiosyncratic property of that lexical item and not a property of the grammar as such. For this reason, we should look to the phrase structure rules for a way to eliminate the redundancy. Moreover, the context sensitivity of the lexical insertion rules is applicable not only to transitive and intransitive verbs, but to verbs with *any* kind of complement and indeed to any category which has complements, i.e. nouns, adjectives and prepositions as well as verbs. In English, the complement always follows the relevant lexical element, as pointed out in the discussion of the head parameter in chapter 2. But the kinds of rewrite rules discussed here cannot capture such generalizations as they are construction specific – i.e. they concern VPs, NPs, APs and PPs separately rather than all phrases. The fact that rule (8) says Verbs precede Noun Phrases in the Verb Phrase tells us nothing about the structure of Noun Phrases.

Chomsky (1970) introduced a structural notation called the X-bar notation which addressed both of these issues. The idea was to take all the lexical material out of the phrase structure rules. When a lexical item was put into a structure, it would bring along all the information from its lexical entry: lexical information would be **projected** into the structure from the lexicon. 'In general, the phrase structure rules expressing head-complement structure can be eliminated apart from order by recourse to a projection principle, which requires lexical properties be represented by categorial structure in syntactic representations: If *claim* takes a clausal complement as a lexical property, then in syntactic representations it must have a clausal complement' (Chomsky, 1986a, p. 82).

The phrase structure rules that are left after this change are very general as they do not need to rely on lexically specific material, such as category and subcategorization information. They also make the general statement that a head of a phrase, the lexical element which projects its categorial nature onto the phrase, precedes its complement, i.e. the element determined by the head's subcategorization frame:

(11) X′ → X YP

This rewrite rule says that a constituent of type X′ is made up of a head, which is also of type X, and its following complement, YP. We will return to the significance of the prime ′, pronounced 'bar', following the X in (11) a little later, but for now take X′ merely to be the phrasal expansion of X. X and Y can then stand for any of the four lexical categories used in the theory, namely N (Noun), V (Verb), A (Adjective) or P (Preposition).

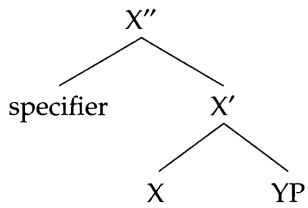
The values for X and Y in any particular structure will depend on which lexical element is inserted into the structure: if X is a transitive verb, then X will be V, and consequently X′ will be V′ (a verb phrase), and Y will be N as transitive verbs have nominal complements in the form of NPs. As the X-bar rule in (11) makes no reference to lexically specific information, this notation eliminates the redundancy of the previous system. It also enables us to make structure-general statements, as anything which is stated about 'X′ is also stated about N, V, A and P simultaneously. So now all we need to do is state that English complements always follow the head X and this statement will be applicable to all nouns, verbs, adjectives and prepositions; rather than having four rewrite rules for four different phrases, one general statement about all phrases suffices. This single powerful rule (11) now includes all the syntactic information that was included in rules (3i) and (3ii), and many others besides regardless of whether an NP or a VP, etc. is concerned.

Chomsky's motivation for the 'bar' part of X-bar Theory was to be able to make further statements that would apply to all constructions in general. The rule in (11) introduces two **projection levels**: structural elements which receive specific properties projected from lexical material. The first level is the head X, also known as the zero level projection or X^0. Above this we have the X′, the first projection of the head. The bar, then, indicates the level of the projection. Chomsky claimed that there is a further projection level required, an X″ (pronounced 'X double bar'). This is introduced by the following rule:

(12) X″ → specifier X′

Thus the X″ contains the X′ and a preceding element known as the specifier of X′. The full structure produced by rules (11) and (12) is as follows:

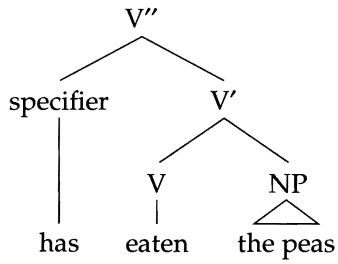
(13)



X″ is the last projection level of the head, equivalent to what we have been referring to as the phrase (XP). As there are no further projections after this, X″ is also called the **maximal projection**.

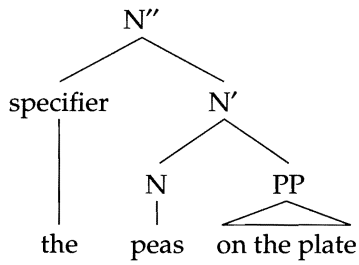
To take some examples of the different types of phrases, the structure captured in rules (11–12) and seen in (13) could be a Verb Phrase:

(14)



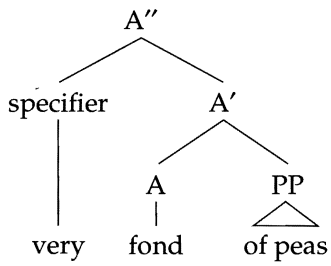
a Noun Phrase:

(15)



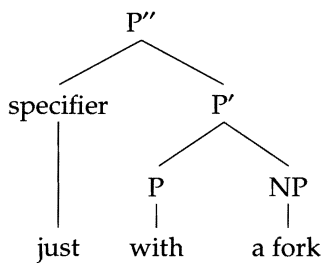
an Adjective Phrase:

(16)



(Note that the triangle is a convention to show that a structure has not been given in full.) Or it could be a Preposition Phrase:

(17)



The specifier position was assumed to be the place for determiner-like elements – determiners in NP, perhaps auxiliary verbs in VP, and degree adverbials in AP:

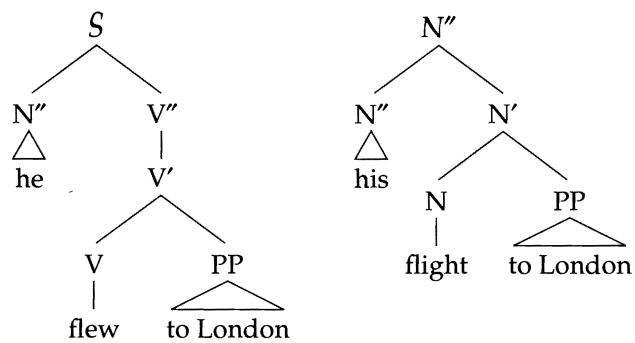
(18) a. *the* slipper
b. *has* slept
c. *too* far

Specifiers differ from complements in that they precede rather than follow the head in English. They are also not subcategorized elements in that they can appear with any head of the relevant type and are not restricted by the head's lexical requirements. Other elements were claimed to occupy the specifier position, notably the possessor in the NP, which is in complementary distribution with the determiner, i.e. you can have one or the other but not both – there is, for instance, never an NP starting *the his* – and hence they seem to occupy the same position:

(19) a. the flight to London
b. his flight to London
c. *the his flight to London

The main point of the structural hierarchy introduced by the X″ is that it allows structural generalizations to extend even further. Compare the two structures below:

(20)



The creation of a specifier position highlights the similarity between the subject of the clause *he* and the possessor of the N *his*, which can be captured in a structural way: both are directly under the maximal node of the structure itself (S in the first case and N″ (NP) in the second) and both are excluded from the rest of the structure (VP/V′ in the first and N′ in the second), which contains the main semantic element of the construction (the verb in the first and the noun in the second). Throughout the 1960s it had often been noted that subjects and possessors behave similarly; introducing the specifier into the theory allowed the statement of general rules which could affect subjects and possessors alike.

During the 1970s, X-bar Theory defined the possible types of rules for the phrase structure part of the grammar: the phrase structure rules were essentially retained from earlier models, subject to the restrictions of X-bar Theory (Jackendoff, 1977). It was still supposed that certain structure-specific facts required structure-specific rewrite rules to account for them, such as the fact that English nouns and adjectives never have NP complements. Thus, the structure of the grammar was believed to be:

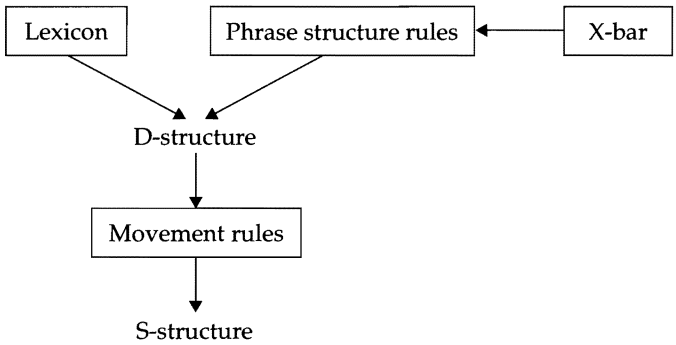
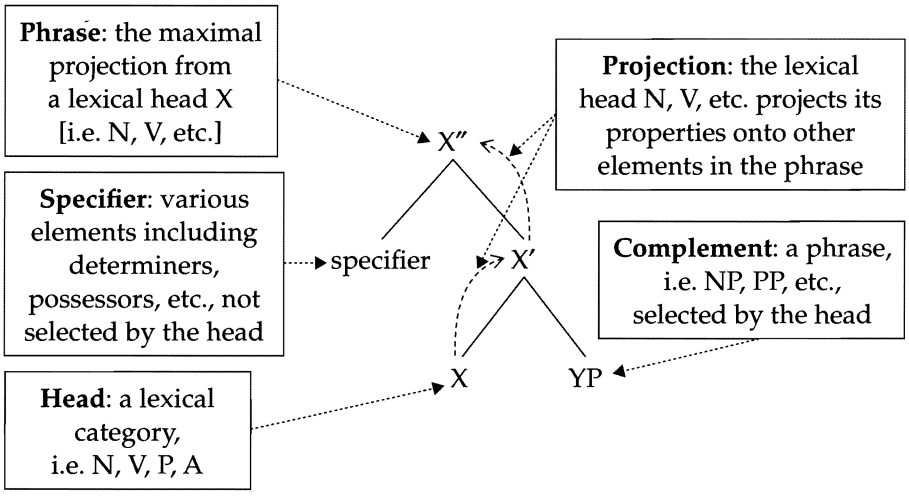


Figure 3.2 Government/Binding Standard Theory (X-bar Theory added)

While GB Theory was to change this view, it nevertheless provides a good example of how the components that were added to the grammar throughout the 1970s restricted the kinds of rules that had previously been envisaged.

RELATIONSHIPS AND CONFIGURATIONS IN X-BAR SYNTAX

This general diagram labels some of the structural relationships that are useful in the discussion of GB Theory.



EXERCISE 3.1

1 Try to represent the following phrases as X-bar syntax trees using the descriptions made so far, identifying the head and the specifiers and complements if present.

the coat with a missing button	too good to miss
offended many people	has destroyed the immune system
his grasp of politics	very interesting
with grace	scatter to the winds

What problems did you find?

2 Identify the heads, their complements and specifiers in the following phrases:

John's picture of Mary	was thinking about it
so fond of a soft bed	too certain that he will win
right on the top shelf	several cups of tea

3.2.2 *Transformational rules and their restrictions*

A similar development happened to the transformational rules in the grammar. During the 1960s few restrictions were placed on what could be a possible transformation, other than what was demanded from empirical considerations and how the description of these empirical facts could be best distributed between the existing components of the grammar. However, once again this led to a situation in which the transformational component contained language- and construction-specific rules so that it was impossible to make generalizations that applied to all transformations.

For example, one idea that emerged was that all movements are short, introduced as the Locality Principle in chapter 2. 'Transformations cannot move a phrase "too far" in a well-defined sense' (Chomsky, 1986a, p. 72). Let us look further at how the English wh-elements such as *who, why* or *what* behave in interrogative constructions. These elements move to the front of the interrogative clause and hence we may conceive of an interrogative wh-movement rule:

(21) D-structure: I asked – Mary likes who.
S-structure: I asked who Mary likes –.

For the time being we will ignore the issue of where the wh-element moves to in order to concentrate on the length of the movement itself. The movement demonstrated in (21) moves the wh-element from its D-structure position as object of the verb to a position at the front of the embedded clause.

(22) D-structure: I asked – Mary likes who.

The next example shows an apparently longer movement:

(23) D-structure: I asked [– Bill thinks [Mary likes who]].
S-structure: I asked [who Bill thinks [Mary likes –]].

i.e.

(24) I asked [– Bill thinks [Mary likes who]].

Here the wh-element is moved out of two embedded clauses.

In fact, it appears that a wh-element can be extracted out of any number of embedded clauses and hence that in principle the interrogative wh-movement rule is unrestricted in terms of how far it can move wh-elements. But this turns out to be false, as the ungrammaticality of the following example shows:

(25) * I asked who Bill wondered why Mary likes –.

This observation was first made by Ross (1967), who accounted for it by introducing a constraint on transformations that prevented a wh-element being extracted out of a clause which already had a wh-element moved to its initial position. This constraint Ross called the **wh-Island Constraint**: clauses which start with wh-elements are 'islands' on which other elements are stranded.

There is a more general way to view this phenomenon, however, which assumes that movement is *always* short, as we argued in chapter 2: movement is short hops, not mammoth leaps. Note that the difference between the grammatical movement in (24) and the ungrammatical movement in (25) is that the position at the front of the second embedded clause is vacant in the first but occupied in the second. If the grammatical case involves not one long movement, but two short ones which makes use of the vacant wh-position, this accounts for the ungrammaticality of (25) by claiming that long movements are not possible:

(26) I asked who Bill thinks – Mary likes –.

I asked who Bill wondered why Mary likes –.

Thus, the reason why clauses that start with a wh-element are islands is that the wh-element which sits in the initial position prevents all other wh-elements from moving to this position and, given that all movements are short, no other wh-element can move out of the clause. The short hop is prevented because there is nowhere for it to land.

The advantage of the short movement account over islands is that the former explains other observations as well. For example, **raising** takes a subject of a lower clause and moves it to the subject position of a higher clause:

(27) D-structure: – seems [John to like Mary].
S-structure: John seems [– to like Mary].

Raising exhibits similar effects to wh-movement: the subject can be moved over long distances only if all intervening subject positions are available to be moved into. If an intervening position is filled, however, long-distance movement is impossible:

(28) John seems – to be certain – to like Mary.

* John seems that it is certain – to like Mary.

The nature of this condition will be detailed in chapter 4. For now the important point to note is that there seems to be a general condition that movements are short and as such we can envisage a general condition which applies to all constraints, which is part of a separate module known as **Bounding Theory**. 'Bounding theory poses locality conditions on certain processes and related items' (Chomsky, 1981a, p. 5). This has a similar relationship to the transformational component to that which X-bar Theory has to the phrase structure component, as seen in figure 3.3.

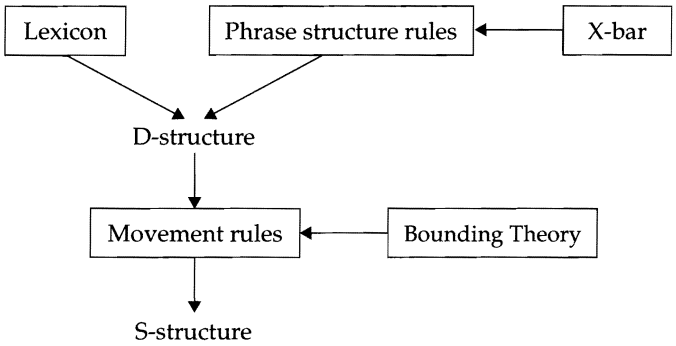


Figure 3.3 Government/Binding Standard Theory (Bounding Theory added)

Just as with X-bar Theory, the introduction of Bounding Theory allowed the grammar to become more general and simple as movement rules could now be stated in far more general ways. During the 1970s this component of the grammar became much reduced, containing very general rules for moving elements such as interrogative pronouns and NPs which were not construction specific (Chomsky, 1977).

The general point of this section has been how the grammar was broken down into modules in order to achieve generality and simplicity within each specific module. In GB Theory this process continued as more modules were introduced, partly to extend the grammar's coverage of grammatical phenomena and partly to simplify the approach to issues already addressed. The next sections will provide more details concerning the modules of the grammar that are specific to D-structure.

BOUNDING THEORY

A generalized theory of restrictions placed on movements. Its principles (to be discussed in the next chapter) ensure that all movements are short. In this way it accounts for why certain constructions (= **islands**) are impossible to move out of, as these would involve movements longer than allowed:

* who did you ask [why [Mary kissed –]]

This sentence demonstrates a **wh-island**, a clause beginning with a wh-element. The ungrammaticality is due to the movement of the object *who* from the wh-island. Bounding Theory explains this by claiming that the movement is too long, with the allowable shorter movement being blocked as there is already a wh-element (*why*) in the relevant position.

Bounding Theory allows transformational rules themselves to be simplified by extracting from them restrictions that would be complicated when stated with respect to particular movements, but are simplified when stated as general conditions on all movements.

EXERCISE 3.2

Identify the moved elements in the following sentences:

He had been watched.
Will that fit through the door?
A man who no one knew arrived at the party.
This suggestion, we will consider next week.
A picture was found of the suspect at the scene of the crime.
Never had they seen such a performance.

3.3 X-bar Theory in Government and Binding

A main consequences of the modularization of the grammar discussed above was that each module could be made simple and general, partly as a result of modules dealing with widespread phenomena rather than specific constructions and partly as a result of extracting complexities from one module and relocating them more generally in a separate module of their own. One of the major simplifications in GB Theory concerned the status of X-bar Theory. As we saw above, X-bar Theory was taken to be a constraining module acting on the phrase structure component of the grammar. In GB Theory, X-bar Theory replaced the phrase structure component altogether, and so took on the role of a constraint on actual structures rather than on the rules responsible for those structures. To see how this is possible, we need to see why the X-bar schema was originally viewed as a constraint.

Although X-bar Theory was designed to capture generalities seen in cross-categorial structures, say the fact that the English head always precedes its complement (the head parameter from chapter 2), certain facts nevertheless seem to be specific to particular phrases. For example, while verbs and prepositions can appear with NP complements, nouns and adjectives cannot:

(29) chew the slipper
in the kennel
* the picture the dog
* fond the dog

This is something that X-bar Theory does not predict as its general rule for introducing complements simply states that a head of *any* category can be followed by a complement of *any* category:

(30) $X' \rightarrow X\ YP$

The specific heads that can appear with specific complements are a matter of lexical information. Yet it cannot just be lexical information that all nouns and adjectives do not take NP complements, as this is not the idiosyncratic behaviour of individual words but something far more general about structures. Thus the need was felt to maintain phrase structure rules to capture generalizations which hold across categories. X-bar Theory then had to be included in conjunction with phrase structure rules.

This was obviously not an optimal solution because of the amount of redundancy between phrase structure rules and X-bar rules. Furthermore it introduces an extra level of complexity to the descriptive content of the grammar: lexical information describes idiosyncratic facts about individual words, phrase structure rules describe facts specific to syntactic categories and X-bar rules describe facts which are general across languages.

Yet the fact that nouns and adjectives do not take NP complements seems to be as much a fact about the distribution of NPs as it does about the complement-taking abilities of nouns and adjectives. Indeed, it is generally accepted that adjectives and nouns *can* select for NP complements as a lexical property, but that these can never surface as bare NPs as they have to undergo a transformation which inserts the preposition *of* in front of them:

(31) the picture the dog → the picture **of** the dog
fond the dog → fond **of** the dog

In GB Theory these facts were ascribed to an entirely different module of the grammar, dealing with the phenomenon of **Case**. More will be said about this in the following chapter, but we need to sketch an outline here. In many languages Case concerns the form that nominal elements bear which realizes information about the semantic or grammatical role of the nominal element. Latin for example has different forms of the noun for different cases; *rex* is the Nominative form for the Subject of the sentence, *regem* is the Accusative form for the Object, *regis*

is the Genitive form to show possession, etc. While Old English used to have many cases, modern English only shows case in the pronouns, *he, him, his* etc.

Case is then associated with certain structural relationships in the sentence; certain positions require the categories that fill them to be in a particular Case. Often subjects are in the **Nominative** Case while objects are in the **Accusative** Case. Case is not always obvious from the surface structure; as pointed out above, it is only visible in English through the forms of the pronouns, unlike languages like Latin or Finnish that make it visible in nouns throughout the sentence. Take:

(32) He saw him.

He is the Nominative Case form of the pronoun and is associated only with the subject of finite clauses, that is to say those having tense, number etc. as opposed to non-finite clauses that lack these features. *Him* is the Accusative Case form and, although this has other uses, its main function is to mark the object of verbs. The difference between the two is the structural position they occupy in the sentence. In GB Theory, following fairly traditional assumptions, the form of the object is 'governed' by the verb: verbs *assign* an Accusative Case to their objects. Prepositions also assign Accusative Case to their objects; therefore any NP complement of a preposition or verb will bear Accusative Case, for instance:

(33) I gave the book to <u>him</u>.

Unlike verbs and prepositions, it can be assumed that nouns and adjectives do not assign any Case to their complements and hence the complement of a noun or an adjective is a Caseless position. Finally, a general principle of Case Theory, called the **Case Filter**, claims that all NPs have to occupy Case positions. Thus no NP can surface in the complement position of a noun or an adjective, even if, as a lexical property, individual nouns and adjectives select for NP complements. However, the insertion of the preposition *of* provides a way of getting round the Case Filter: the preposition assigns the missing Case and renders the position a Case position. That is, in:

(34) fond him

him cannot have Case because it is in the complement position of the adjectival phrase. Inserting an *of*:

(35) fond of him

allows *him* to be in the accusative case because it is now governed by a preposition.

Facts about the surface distribution of NPs therefore fall out from the Case module of the grammar and do not have to be stated as special rules in the phrase structure component. Once all category-specific facts have been accounted for in this way, the phrase structure rules themselves are no longer needed and only X-bar rules and lexical information remain.

The grammar can therefore be simplified as in figure 3.4.

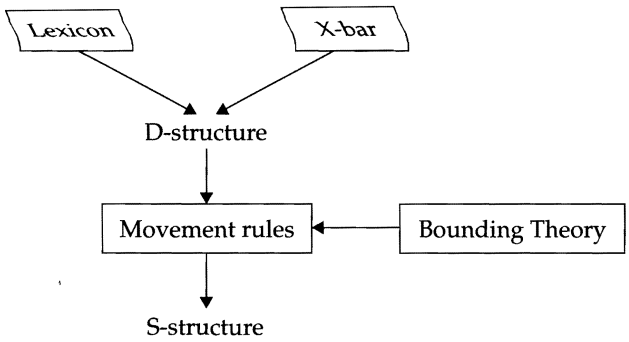


Figure 3.4 Government/Binding Theory (X-bar replacing phrase structure)

CASE THEORY

A module of the grammar that determines the distribution of NPs through a requirement that all NPs must be in Case positions at S-structure, known as the **Case Filter**. Case positions are those **governed** by certain Case assignors:

- Verbs and Prepositions govern and Case-mark objects as **accusative**:

  e.g. John saw *him*.
  John showed the picture to *him*.

- Nouns and adjectives do not Case-mark their objects and hence they do not have bare NP complements at S-structure.

However, by insertion of the pleonastic preposition *of*, the NP complements of nouns and adjectives are allowed to surface:

e.g. * a picture *him* a picture *of him*
* fond *him* fond *of him* etc.

Completely replacing phrase structure rules with X-bar principles gets rid of the redundancy of categorial information being stated both in the lexicon and in the phrase structure component. Under this view, the X-bar principles regulate a category-neutral structure and categorial information enters as lexical items are inserted. In this way it is the lexicon that determines the specific properties of actual phrases through the notion of projection. We can imagine a process in which we first construct an X-bar structure, such as the following:

(36)
```
X″
|
X′
|
X
```

We then insert a lexical item into the head position:

(37)
```
X″
|
X′
|
X
|
sleep
```

As this lexical item is a verb, its verbal category will be projected first to the X node, making it a V:

(38)
```
X″
|
X′
|
V
|
sleep
```
projection

As X is the head of X′, the projection will continue to this node, making it a V′:

(39)
```
X″
|
V′
|
V
|
sleep
```
projection

Finally, as X′ is the direct head of X″, the projection will continue further up to the V″:

(40)
```
V″ (VP)
|
V′
|
V
|
sleep
```
projection

As this is the maximal projection, the projection of the verbal category ends here; other lexical items inserted into the larger structure will determine other aspects of the sentence as a whole. It is clearly important that structures be regulated by lexical information: categorial features projected into a structure must be anchored to the properties of the inserted lexical items. 'An X-bar structure is composed of projections of heads selected from the lexicon' (Chomsky, 1993, p. 8). Thus a general principle governing the relationship between inserted lexical elements and features projected into the structure is needed, called the **Projection Principle**:

(41) **Projection Principle**
The categorial properties of structures are projected from the lexicon.

Unfortunately the Projection approach tends to produce cumbersome trees that spread across the page since every phrase is projected upwards maximally, losing the clear advantages of the visual presentation of structures as trees. The Minimalist Program (MP) approach to structure using a process called Merge, in which the trees are built up out of pairs of categories, tends to produce a starker tree and will be outlined in chapter 7.

THE PROJECTION PRINCIPLE

A principle which ensures that lexical information remains constant at all levels of syntactic representations (e.g. D-structure and S-structure). Lexical information enters at D-structure with the insertion of lexical elements. This information is then projected into the structure in accordance with X-bar principles.

The next chapter shows how the Projection Principle also serves to prevent this information from being changed by the action of transformations.

Some elements of structure, however, cannot be analysed as complements or specifiers but can nonetheless accompany the head within a phrase. Consider the following examples:

(42) a. John slept.
b. John slept very soundly.

The verb *sleep* is intransitive and hence does not select for a complement, so the adverbial phrase *very soundly* in (42b) cannot be a complement. This element is known as an **adjunct**. Adjuncts are non-selected modifiers of heads and as such they can appear fairly freely, being both optional and able to be included in a structure in indefinite numbers (unlike complements). Thus, in addition to the examples in (42), we can also have:

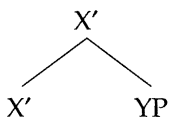
(43) John slept very soundly in his bed all night dreaming beautiful dreams.

There has been a certain amount of controversy about how adjuncts are to be incorporated into a structure within the X-bar framework. One idea is that they are introduced by a **recursive** rule, which introduces another instance of the projection level it expands and thus may apply again to its own output, creating a loop that can go on indefinitely. In general recursion is a property of a rule that can call on itself; Chomsky sees it as one of the unique capabilities of the language faculty (Hauser et al., 2002). An instance of a recursive rule might look like the following:

(44) X′ → X′ YP

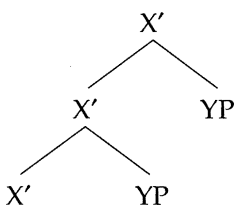
This rule produces the following kind of structure:

(45)



Now the lower X′ could in turn be expanded into an X′ recursively, yielding:

(46)



Obviously this could go on for ever and hence an indefinite number of adjuncts could be added to the structure, which accords with adjunct properties, as seen by the limitless recursion of *very* in English:

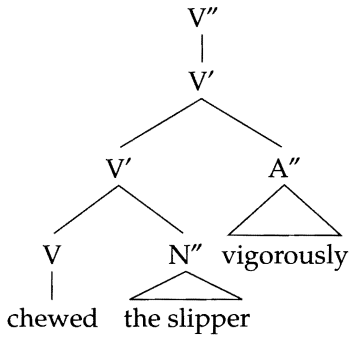
(47) The prime minister was very very very very . . . wrong.

Furthermore, if adjuncts are inserted by this kind of rule, we predict that they should be further from the head than are complements, as complements are sisters to the head, i.e. alongside it in the tree beneath the same node. In many cases this prediction is borne out:

(48) a. The dog chewed the slipper vigorously.
b. * The dog chewed vigorously the slipper.

The VP structure in (48a) would therefore look like the following:

(49)



However, this treatment of adjuncts is not without problems, the main one being the redundancy introduced by the structural approach to the recursive nature of adjuncts. Given that adjuncts are not selected elements, their inclusion into a structure is predicted to be unrestricted on lexical grounds; it is therefore redundant to have this fact follow from the mechanisms which regulate the structure. We will return to the treatment of adjuncts later when we will be in a better position to appreciate alternatives.

3.4 Theta Theory

The view of D-structure that developed out of the 1960s was as a structural representation of certain semantic relationships between elements. 'Phrase structure rules of a very simple kind generate an infinite class of D-structures that express semantically relevant grammatical functions and relations' (Chomsky, 1981a, p. 67). To take the example of the passive structure, the element that sits in the subject position at S-structure is interpreted as the object of the verb and hence is assumed to sit in object position at D-structure:

(50) D-structure: – was chewed the slipper.
S-structure: The slipper was chewed –.

However, this raises questions about the definition of object and subject positions and why certain elements sit in them but not others. When we compare subjects and objects, one of the most obvious facts is their relationships with the verb. In:

(51) The dog chewed the slipper.

the dog is interpreted as the one doing the chewing and *the slipper* is the thing getting chewed. This is not because of our pragmatic knowledge – knowing that dogs tend to chew things and slippers are the likely objects of a dog's attentions. If the subject and object are switched round, we get:

(52) The slipper chewed the dog.

This sounds a pragmatically anomalous sentence in which the slipper is doing the chewing and the dog getting chewed. Yet people can still interpret the sentence even if they find it ridiculous. This means that the subject and object positions are associated with certain interpretations that are forced by the syntax.

From a semantic point of view, what is involved here is the relationship between elements known as **arguments** and **predicates**. A predicate is something which expresses a state or a relationship and an argument is something that plays a role in that state or relationship. Thus in (52) the predicate is the verb *chew* expressing a relationship and the arguments are the NPs *the dog* and *the slipper*, the things involved in the relationship. Different arguments play different roles with respect to the predicate. Thus, the subject of *chew* plays the role of the 'chewer' and the object plays the role of the 'chewee', the thing chewed.

More generally the subject of a large number of verbs is the one that deliberately and consciously carries out the action described by the verb, a semantic role known as **agent**. *The dog* is then the agent in (52). The argument that is acted upon by the agent, i.e. the one sitting in object position, is called the **patient**, exemplified by *the slipper* in (52). Roles such as agent and patient are known generally as **thematic roles**, or **θ-roles (theta roles)** for short. To take some examples, in:

(53) Pete drank a pint of Adnams.

Pete is the agent in subject position, who deliberately drank something; *a pint of Adnams* is the patient in object position that the agent drank. In:

(54) The government banned speeches fomenting terrorism.

the subject *the government* is the agent that banned something; the object *speeches fomenting terrorism* is the patient that the agent banned.

However, not all subjects are interpreted as agents, and not all objects as patients. For example in:

(55) John sent a letter to Mary.

John is the agent of *send* and *a letter* is patient. *Mary* has the role of **recipient**, the receiver of something, indicated by the preposition *to*. Alternatively, we might think of *Mary* as the **goal**, i.e. the end point of the action described by the verb. In:

(56) Mary received a letter from John.

even though *Mary* is subject, this element is still interpreted as recipient and certainly not as the agent: Mary is not the one who deliberately and consciously performed the act of receiving – you can rarely choose to receive, something e-mail users will bear witness to. Next, consider the following:

(57) a. The dog chewed the slipper.
b. The dog saw the slipper.

As we have said, the object *the slipper* is interpreted as patient in (57a) but not in (57b) where nothing actually happens to the slipper as a result of the dog seeing it. We might call this semantic role **theme**. Moreover, the subject of *see* in (57b) is not an agent as there is no action performed in this case. We call this argument type an **experiencer**. Clearly the semantic role that an argument bears depends on the predicate: the subject of *chew* is an agent and its object is a patient, while the subject of *receive* is a recipient and the object of *see* is a theme. This must then be lexical information about how individual verbs behave, showing that *chew* is different from *see*, and so it must be stored in the lexical entry for each predicate:

(58) *chew*: [__ NP], <agent, patient>
receive: [__ NP PP], <recipient, theme, source>
see: [__ NP], <experiencer, theme>

These lexical entries therefore not only include the subcategorization frames of the verbs detailing what complements they take but also a **theta grid** supplying information about the roles that their arguments take.

Although we have now determined that different predicates take different types of arguments, we have yet to say how this affects the specific interpretation of particular arguments in particular positions. In GB Theory it is claimed that predicates **assign** their θ-roles to specific structural positions and that any element sitting in those positions will be interpreted as bearing the assigned roles. So the agent of *chew* is assigned to the subject position and the patient is assigned to the object position:

(59) The dog chewed the slipper.
agent patient

The verb *chew* has then affected the structure of the sentence by assigning its two θ-roles to the appropriate positions.

We can distinguish two types of θ-role, as seen in:

(60) a. John broke the window.
b. The hammer broke the window.

The subject of *break* in fact can have several possible interpretations: either as agent, as in the more natural interpretation of (60a), or as **instrument**, i.e. something through which an action is carried out, as in (60b). There is clearly some interaction between the role the subject is interpreted as bearing and the subject itself, as hammers are not the sort of thing that can be agents (except of course in a metaphorical sense *The hammer of God*), though *John* could be interpreted as instrument in the case that someone throws him through the window! However, the interpretation of the object remains the same in both the sentences in (60): *the*

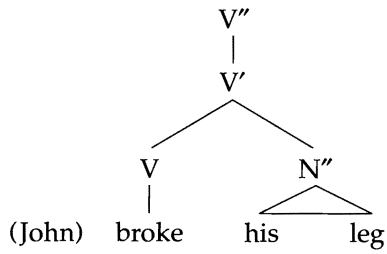
window is patient no matter what choice of subject we make. This suggests that the interpretation of the object is solely determined by the verb itself, contrasting strikingly with the interpretation of the subject, as seen in:

(61) a. John broke the window.
b. John broke his leg.

In (61a) the subject can be interpreted as agent or, less naturally, instrument. But in (61b) the subject has a different set of possible interpretations. In the most natural interpretation *John* is seen as the one whose leg is broken rather than the one that actually does something, though the agent interpretation is possible: John could deliberately break his own or, perhaps more likely, someone else's leg. The instrument interpretation is not available in this instance. Thus the range of interpretations differs in both cases, showing that the interpretation of the subject is determined not solely by the verb, but also by the choice of the object. For these reasons the object is known as the **internal argument**, suggesting that it is closer to the verb both semantically and structurally, while the subject is known as the **external argument**, suggesting a greater distance between it and the verb.

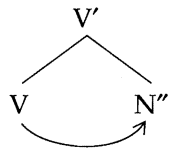
We can now consider the principles that govern θ-role assignment. The internal argument is always the complement of the verb and structurally speaking is the verb's **sister**, i.e. is one of two constituents immediately dominated by the V′, their **mother**:

(62)



The internal θ-role is therefore assigned to the sister of the assigning head. This not only explains why the internal argument is close to the head that θ-assigns it, but also accounts for the fact that only the verb determines the range of θ-roles available to the object: inside the V′ there is nothing but the verb to influence the internal argument:

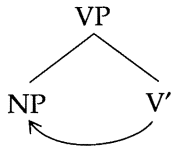
(63)



The case of the external argument is different. In particular, both the verb and its complement play a role in determining the θ-roles available to the subject. This suggests that what actually assigns the external θ-role is the combination of

the head and its complement: in other words, the V′. If a similar restriction to that for the internal θ-role applies to the assignment of the external θ-role, we might expect it to be assigned to the sister of the V′, which is the specifier position:

(64)



Towards the end of the 1980s, it became increasingly popular to hypothesize that the subject actually originates inside the VP – the **VP-Internal Subject Hypothesis**. Koopman and Sportiche (1991) argued extensively in favour of this on both theoretical and empirical grounds. For now the proposal will be accepted on the grounds that it simplifies the principles governing θ-role assignment.

The general principle governing θ-role assignment is therefore that θ-roles are assigned to the sister positions of the θ-role assigning element, known as the **Sisterhood Condition**:

(65) **Sisterhood Condition**
A θ-role assigning element assigns its θ-role to its sister.

'θ-marking meets a condition of "sisterhood" that is expressible in terms of X-bar theory . . . : a zero-level category α *directly θ-marks* β only if β is the complement of α in the sense of X-bar theory' (Chomsky, 1986b, p. 13). This condition, along with restrictions asserted by X-bar Theory, plays a role in the definition of a well-formed D-structure, putting the relevant arguments into appropriate structural positions. The Sisterhood Condition is one of the principles of a module called **Theta Theory (θ-Theory)** which applies directly to D-structure:

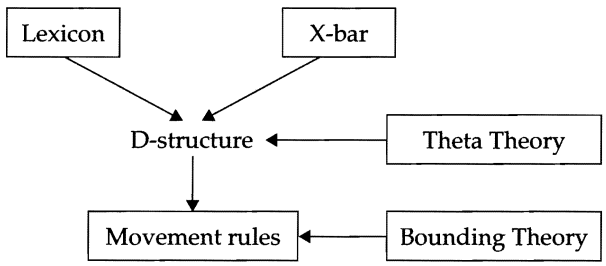


Figure 3.5 Government/Binding Theory (Theta Theory added)

Another principle of Theta Theory regulates how θ-roles are assigned. For example, an argument cannot be inserted into a structure without having a legitimate θ-role. If there is an object with an intransitive verb, the result is ungrammatical:

(66) * Mary smiled John.

(apart from a few exceptions where the object reiterates the verb *She smiled a big smile*). In this case, *John* receives no θ-role and the ungrammaticality indicates that arguments must bear θ-roles. Moreover, arguments are not able to distribute a single θ-role over several positions:

(67) * the dog chewed the slipper the bone

The conclusion is then that every argument must have at least one θ-role assigned to it and that every θ-role can be assigned to one argument at most. Indeed, it appears that an argument can only bear one θ-role:

(68) * the dog chewed the slipper disintegrated

The slipper cannot be both patient of chewing and experiencer of disintegration at the same time, without adding a substantially more complex structure.

This suggests that θ-roles and arguments are in a one-to-one correspondence with each other, captured by a principle known as the **Theta Criterion (θ-Criterion)**:

(69) **Theta Criterion**
All theta roles must be assigned to one and only one argument.
All arguments must bear one and only one theta role.

Together, the Sisterhood Condition and the Theta Criterion constitute Theta Theory.

THETA THEORY

A module of the grammar dealing with the assignment of semantic roles (θ-roles), such as agent, patient and goal, to arguments in a sentence. It consists of two basic principles:

- The **Sisterhood Condition** states that θ-roles are assigned to sisters of the assigning element. **Internal θ-roles** are assigned directly from a head to its complement, while **external θ-roles** are assigned compositionally by the head and its complement, via the X′ to its sister, the specifier. This assumes the **VP-Internal Subject Hypothesis**, which claims that the subject originates in the specifier of the VP and then moves to the specifier of IP at S-structure.
- The **Theta Criterion** states that θ-roles can be assigned to only one argument and arguments can bear only one θ-role. Thus θ-roles and arguments are in a one-to-one correspondence.

EXERCISE 3.3

1 Devise a sentence for each of the theta roles that have been mentioned in this section and then devise a single sentence that exemplifies all or as many as possible of them:

patient	recipient
experiencer	source
agent	theme

2 The subject of a verb such as *seem* may be realized by the semantically empty pronoun *it*:

It seems John likes fishing.

However, there is another way to express the meaning of this sentence without using a meaningless subject:

John seems to like fishing.

How can this sentence mean virtually the same thing as the one with the meaningless subject if θ-roles are assigned to sister positions?

3.5 Control Theory and null subjects

So far we have been concentrating on the structure of phrases and have said relatively little about the higher structure of the sentence. The next section will investigate sentence structures in more detail, but certain aspects of the treatment of sentential elements, particularly the subject, need to be introduced first.

Consider the following sentence:

(70) There arrived a mysterious package.

Because of the Theta Criterion that arguments and θ-roles are in a one-to-one correspondence, the verb *arrive* takes a single argument: the one who arrives. In (70) this is obviously *the mysterious package*, which bears the θ-role assigned by *arrive*. This means that the subject of this sentence, *there*, appears to lack a θ-role. Indeed, this element is essentially meaningless and is referred to as an **expletive** or **pleonastic** subject, both terms meaning 'meaningless' in this context. So the subject *there* does not actually violate the Theta Criterion, as the subject is not an argument of any predicate and so does not need a θ-role. However, one might wonder what it is doing there in the subject position at all. Indeed the subject is obligatory in English; leaving it out is unthinkable:

(71) * Arrived a mysterious package.

Since virtually all sentences of English have subjects, this suggests strongly that a grammatical principle is involved, namely the **Extended Projection Principle (EPP)**, which states that all clauses must have a subject. 'The two principles – the projection principle and the requirement that clauses have subjects – constitute what is called the extended projection principle (EPP)' (Chomsky, 1986a, p. 116). Thus, in situations where there is no semantic subject, an expletive subject has to be inserted, as in (70).

The EPP does, however, appear to have several exceptions. One typically involves the subject position of non-finite clauses, which appear to be subjectless in most languages:

(72) The dog tried [to chew the slipper].

Although there is no apparent syntactic subject of the non-finite clause enclosed in brackets in (72) *[to chew the slipper]*, semantically the subject of the main clause *the dog* clearly acts as the missing subject: the dog is the one both doing the trying *and* the chewing. However, allowing both verbs to assign their external θ-roles to this subject would mean that one argument would end up with two θ-roles, in violation of the Theta Criterion. Nor can the subject be the sister of two V's at the same time, and so it would not satisfy the Sisterhood Condition either.

Semantically, the situation is very similar to the following:

(73) The dog thinks [he chewed the slipper].

One obvious interpretation of this sentence has the dog doing both the thinking and the chewing. But this would involve two distinct arguments: *the dog* and *he*. We can interpret both of these arguments as the same because the pronoun *he* may be referentially dependent on the higher subject *the dog*. The situation where two elements are *co-referential* is often indicated by giving them the same *index*, shown in subscripts:

(74) The dog_i thinks [he_i chewed the slipper].

The problem of the subjectless non-finite clause could be solved if we were to suppose that there is a pronoun-like element which sits in the subject position and which is referentially dependent on the higher subject:

(75) The dog_i tried [$pronoun_i$ to chew the slipper].

Under this assumption, the non-finite clause has a subject, thus satisfying the EPP and there are, moreover, two independent subjects to bear one θ-role each, satisfying the Theta Criterion. Of course, the problem with this suggestion is that no such pronoun is visible in the actual sentence. But, put another way, perhaps the situation does not sound so hopeless. What sentence (75) claims is that there is a syntactic and semantic subject of the non-finite clause, but that it is not phonetically realized. It is not new to talk about there being phonologically empty syntactic elements. Most grammarians accept the existence of null morphemes which

fill out paradigms to keep the description as regular as possible. For example, Hungarian intransitive verbs show the following conjugations:

(76)	sétálok	I walk
	sétálsz	you (sing.) walk
	sétál	he/she/it walks
	sétálunk	we walk
	sétáltok	you (pl.) walk
	sétálnak	they walk

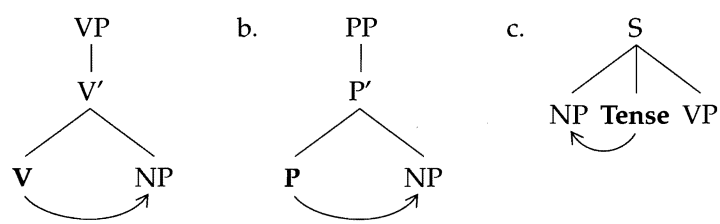
Only when the subject is third person and singular *sétál* does the verb have no inflection. However, the alternative to assuming the absence of any inflection is to assume that there is actually a morpheme here which is unpronounced. The same could be claimed for the null pronoun subject of the non-finite clause: it too is simply unpronounced.

The null pronoun subject of non-finite clauses is often called **PRO** and has a number of properties peculiar to itself. First, while it is obviously a nominal element, it has a far more restricted distribution than other NPs. PRO is only ever found in subject position in non-finite clauses and is banned from object position and finite clause subject position:

(77) a. * The dog chewed PRO.
b. * The dog thinks [PRO chewed the slipper].

These positions are said to be **governed**, which is a crucial technical relationship in GB Theory. Essentially a governed position has a relationship to a particular element called a **governor**. Lexical heads, such as nouns, verbs, adjectives and prepositions, are governors along with the inflection of the finite clause. The former all govern their complement positions whilst the latter governs the subject position. To take some examples:

(78) a. b. c.



In (78), the bold heads are all governors and the arrows show the elements that they govern. As PRO is unable to occupy any of these positions, it appears that it cannot sit in a governed position and, since nothing governs the subject position of the non-finite clause, the non-finite marker *to* not being a governor, this is the only position in which this element can occur. This is known as the **PRO Theorem**:

(79) **PRO Theorem**
PRO can only sit in ungoverned positions.

This theorem will figure in the next chapter, where an attempt to derive it from more basic principles will be discussed.

The referential capabilities of PRO also distinguish it from other elements. As seen in (74), a pronoun like *he* can be referentially dependent on another element in a higher clause. However, clearly this is not necessary and the pronoun could refer to someone else. Hence the following representation, where the fact that the NP and the pronoun have different indices ($_{i\,j}$) indicates disjoint reference, is perfectly possible:

(80) The dog_i thinks [he_j chewed the slipper].

i.e. it was someone unnamed such as the owner that chewed the slipper.

PRO, on the other hand, is less flexible and often its reference is fixed to a particular element. For example consider the following sentences:

(81) a. John persuaded the dog_i [PRO_i to stop chewing his slipper].
b. $John_i$ promised the dog [PRO_i to stop chewing his slipper].

In the first example, it is *the dog* who will stop chewing the slipper, but in the second it is *John* who will desist from slipper-chewing! The element which fixes the reference of PRO is the **controller**. 'Control theory determines the potential for reference of the abstract pronominal element PRO' (Chomsky, 1981a, p. 6).

The determination of which element is the controller is complex. As can be seen in examples like (81), the properties of the verb which selects the non-finite clause as a complement have some part to play. The verb *persuade* is referred to as an **object control** verb since its object acts as the controller. The verb *promise*, on the other hand, is a **subject control** verb since its subject is the controller.

Further complications arise with other examples:

(82) a. John asked the dog [PRO to stop chewing his slipper].
b. John asked the doctor [how PRO to stop chewing his slipper].

As (82a) shows, *ask* appears to be an object control verb. However, in (82b) it is not the object which controls PRO. In this case, the subject could be the controller (i.e. John wants to know how he can stop chewing slippers) or there might not be any controller at all (i.e. John wants to know how it is possible for anyone to stop chewing slippers). In the second instance, PRO has **arbitrary reference**, similar to the generic pronoun *one*:

(83) John asked the doctor how one can stop chewing slippers.

Finally in:

(84) [PRO to be or not to be], that is the question.

we see another instance of PRO with arbitrary reference; being or not being is not ascribed to anyone in particular.

The principles governing control phenomena are clearly complex. They form another module of the grammar, known as **Control Theory**, which also applies to D-structure. This can now be added to our model:

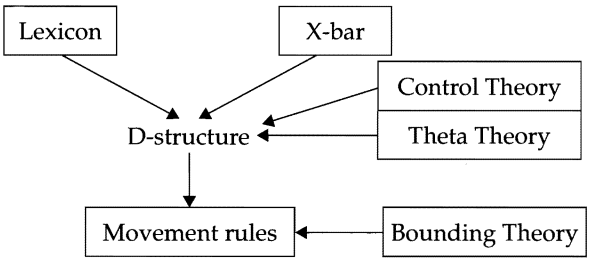


Figure 3.6 Government/Binding Theory (Control Theory added)

CONTROL THEORY

A module concerning the reference of the empty subject PRO, which has either **controlled reference** or **arbitrary reference**. In controlled cases, the controller is either the subject or the object, depending on the verb:

John promised Bill PRO to leave.
John persuaded **Bill** PRO to leave.

In arbitrary cases, PRO is interpreted as having *generic* reference, similar to the reference of the pronoun *one*:

PRO to leave would be impolite (for one to leave would be impolite).

There is yet another instance where the EPP appears not to hold, though this phenomenon is more restricted cross-linguistically than control phenomena. In some languages a pronoun subject of a finite clause can be 'dropped'. For example, the Hungarian verbs in (85) can all be taken as complete sentences whether or not there is an overt subject:

(85) (Én) leülök.
I sit down.
(Te) leülsz.
You sit down.
(Ő) leül.
He/she/it sits down.
(Mi) leülünk.
We sit down.
(Ti) leültök.
You (pl.) sit down.
(Ők) leülnek.
They sit down.

This possibility is not available for all languages, as can be seen from the following English verbs, which by themselves cannot constitute a complete sentence:

(86) * (I) sit down.
* (He) sits down.

Once again, however, the fact that these subjectless sentences are interpreted as though they do have pronominal subjects in some languages indicates another instance of a phonologically null element. The phenomenon is often referred to as 'pro-drop'. Languages which allow such null-subject sentence are called **pro-drop languages** and those which do not are called **non-pro-drop languages**.

The pro-drop parameter seems to be one of the major parameters of variation between languages, sometimes cutting across language families. For example French is a non-pro-drop language, but all other Romance languages, including Italian, Spanish, Portuguese and Romanian, are pro-drop languages. The Germanic languages are all non-pro-drop while other language families, such as Slavic, are all pro-drop. Indeed across the world the vast majority of languages appear to be pro-drop. The box gives some examples of languages that are pro-drop and non-pro-drop.

Some pro-drop languages (with null subjects, i.e. allowing the empty element *pro* to be subject of the sentence)		**Some non-pro-drop languages** (without null subjects, i.e. not allowing *pro* as subject)
Italian	Chinese	German
Arabic	Greek	French
Portuguese	Spanish	English
Hebrew	Japanese	Dutch

There is some controversy concerning how to analyse pro-drop. On the one hand it is tempting to include it with control phenomena, accounting for the parameterization in terms of greater and lesser restrictions on the distribution of PRO. However, the standard approach is to separate pro-drop and control phenomena as involving two different phonologically empty pronouns. The empty pronoun in pro-drop phenomena goes by the name of 'little pro', i.e. **pro**, as opposed to its big cousin **PRO**. One advantage of having two distinct empty pronouns is that the difference between pro-drop and non-pro-drop languages shows up as presence or absence of *pro*: pro-drop languages have *pro*, non-pro-drop languages do not. Another advantage is that we can more easily describe differences in the behaviour of the two pronouns if they are treated as separate. For example, the referential possibilities of *pro* tend to be very similar to those of an overt personal pronoun, unlike PRO which has controlled referential possibilities.

The analysis of the parametric difference between pro-drop and non-pro-drop languages also has its controversies. One of the first analyses was Rizzi (1982), who claimed that whether a language licenses *pro* in subject position depends on its agreement system. Above we mentioned the notion of government, claiming lexical heads and finite inflection to be governors. Another, stronger notion is **proper government**, which claims there is a set of proper governors restricted to lexical

heads, and excluding the finite inflection. Many pro-drop languages differ from non-pro-drop languages in terms of the richness of their agreement systems. Compare the Hungarian example to the English one:

(87)	1st sing.	sétál-ok	walk
	2nd sing.	sétál-sz	walk
	3rd sing.	sétál	walk-s
	1st pl.	sétál-unk	walk
	2nd pl.	sétál-tok	walk
	3rd pl.	sétál-nak	walk

Hungarian has a different agreement form for each of the six members of the agreement paradigm whereas English only has two forms in all. The richness of the agreement systems of pro-drop languages, Rizzi claimed, allows their finite inflections to be treated as proper governors and therefore the condition which licenses *pro* is that it must be properly governed. This in turn is part of a more general condition to be discussed in the next chapter called the **Empty Category Principle**.

The positive aspect of Rizzi's original analysis of the pro-drop parameter was that the proposed small difference between languages – whether or not finite inflection is a proper governor – accounts for a wider range of phenomena than just whether or not the language allows a pro subject. Amongst other properties, Rizzi listed the following as being typical of pro-drop languages:

- not having pleonastic subjects
- the ability to 'invert' subjects and VPs
- the ability to extract wh-elements more freely out of certain clauses.

'[An] interesting topic . . . is the clustering of properties related to the pro-drop parameter, whatever this turns out to be. In pro-drop languages (e.g. Italian), we find among others the [above listed] clustering of properties. . . . Non-pro-drop languages (e.g. French and English) lack all of these properties, characteristically' (Chomsky, 1981a, p. 240). We can demonstrate these properties with the following Italian examples:

(88) Sembra che Gianni sia ammalato.
(seems that John is ill)
It seems that John is ill.

While the English sentence in (88) needs an expletive subject, there is no such element in the Italian version.

(89) Ha telefonato Gianni.
(has telephoned John)
John has telephoned. (* has phoned John)

In this case the subject follows the VP, which is generally not possible in English.

Finally (90) shows an interrogative clause, which, as the English translation demonstrates, is not possible in English.

(90) Che credi che verrà?
(who believe-2.sing. that come)
* Who do you believe that will come?

The grammatical English version obligatorily leaves out the word *that*, known as a **complementizer**:

(91) Who do you believe will come?

The point is that sentences (88–90) are all grammatical in Italian, a pro-drop language, but are ungrammatical in English, a non-pro-drop language. What is more, the grammaticality always has to do with the subject of a finite clause. In Rizzi's analysis this followed from the assumption that the Italian finite inflection properly governs the subject whereas the English one does not.

In the early days of GB Theory it was hoped that this sort of analysis would provide plenty of explanatory content as it suggested that small grammatical differences could be responsible for quite wide-ranging differences between languages in terms of the surface phenomena that they demonstrate. In turn this could play a role in accounting for how the process of language acquisition could be completed so quickly and so effortlessly by the child. All a child needs to learn is whether the finite inflections of their language are a proper governor or not, and knowledge of a great many other things follows. 'When this parameter is set one way or another, the clustering of properties should follow. The language learner equipped with the theory of UG as part of the initial state requires evidence to fix the parameter and then knows the other properties of the language that follow from this choice of value' (Chomsky, 1981a, p. 241).

Rizzi (1986) modified his analysis slightly to accommodate the fact that in some languages *pro* seems to be allowed in positions other than the subject of a finite clause. For example Hungarian pronominal possessors may be dropped as well as subjects, as in:

(92) Láttam az (te) anyukádat.
(saw-1st sing. the (you) mother-2nd sing.-acc.
I saw your mother.

And in Italian objects can sometimes go missing:

(93) Un dottore serio visita – nudi.
(a doctor professional visits naked)
A professional doctor examines his patients when (they are) naked.

Rizzi claimed that *pro* needs to be licensed by certain elements, the identity of which is open to parametric variation, and moreover its content must be recovered from the licensor. English selects no licensor for *pro* and hence *pro* cannot

appear. Standard pro-drop languages, on the other hand, select finite inflection as a licensor of *pro* and hence they have *pro* subjects of finite clauses. Hungarian also has a nominal licensor and Italian a verbal one. Thus the range of licensors differs from one language to another. As for the recovery of the content of *pro*, a morphologically rich inflection allows this in a straightforward fashion; that is to say you may be able to tell the subject's number and gender from the verb forms even if the subject is actually absent. For instance in the Hungarian sentence:

(94) Lemegyünk a kocsmába.
(down-go the pub-to)

the *ünk* shows the verb is first person plural and that it agrees in number and person with the *pro* subject, which is therefore interpreted as meaning 'we' and the sentence means *We're going down the pub*. Note further that the Hungarian nominal also agrees with the possessor facilitating the recovery of the content of this element. Something special must be assumed for the content of the *pro* object in Italian as it is clear that the content cannot be recovered directly from the verb. The content of the missing object is also special in that it is always generic, similar to the interpretation of the pronoun *one* in the following:

(95) One has to put up with such things.

Building on the fact that verbs assign θ-roles to their objects, Rizzi suggests that a *pro* which is θ-marked by its licensor will receive a generic interpretation. As the subject is not thematically related to the inflection, this will obviously not be true for *pro* in subject position.

This development helps account for further facts that were problematic for the original theory. For example, while German is not normally considered a pro-drop language, it can drop expletive subjects as in:

(96) Gestern wurde *pro* lange diskutiert.
(yesterday was-3.sing. long discussed)
Yesterday it was discussed/the discussion went on until late.

So German finite inflection is a licensor for *pro*, but it is not rich enough to allow referential content to be recovered from it. A non-referential, expletive *pro* has no content and therefore is possible.

However, this theory faces a number of problems. For one thing, there are some non-pro-drop languages which demonstrate similar phenomena to those exemplified in (88–90). Scandinavian languages, for example, allow wh-movement out of clauses which start with a complementizer, as shown by the following Norwegian example:

(97) Hvem tror du at har stjålet sykkelen?
(who think you that has stolen bike-the)
Who do you think (*that) has stolen the bike?

but, like all Germanic languages, they are not generally pro-drop:

(98) * *pro* Glitrer som diamanter.
(glitters like diamonds)
It glitters like diamonds.

Furthermore, some pro-drop languages do not display the same set of effects as Italian. Japanese, for example, is a pro-drop language, but does not demonstrate any overt wh-movement. Worse still, there are pro-drop languages which do not have rich agreement systems. Chinese is a pro-drop language, but has no verbal agreement whatsoever:

(99) Wo (I)
Ni (you sing.)
Ta (he/she/it)
Women (we)
Nimen (you pl.)
Tamen (they)
} ai Lisi.
like Lisi.

Essentially the pro-drop parameter can be set independently of the other phenomena, disappointingly from the point of view of the theory of parameter setting.

Another approach has attempted to explain why languages with rich inflection systems, such as Italian and Hungarian, and languages with no inflections at all, such as Chinese, should be pro-drop, whereas languages with patchy inflection systems, such as English, are not. Comparing the three types of language shows what those with rich inflections and those with no inflections have in common:

(100)

	Italian	Chinese	English
1st sing.	parlo	shuo	speak
2nd sing.	parli	shuo	speak
3rd sing.	parla	shuo	speaks
1st pl.	parliamo	shuo	speak
2nd pl.	parlate	shuo	speak
3rd pl.	parlano	shuo	speak

While Italian has a different form for each part of the present tense paradigm, Chinese has the same form throughout. In English, however, while most parts of the paradigm are the same, one form *speaks* is different from the others. Thus what connects Italian and Chinese is that they both have uniform paradigms: uniformly different or uniformly the same. The English paradigm is non-uniform, with *speaks* being the exception that destroys uniformity of both types. Jaeggli and Safir (1989)

propose that the notion of **morphological uniformity** explains pro-drop and that a language's inflections may be analysed as + or – uniform. A +uniform inflection is what licenses *pro*.

Although this theory explains why languages at the opposite ends of the inflectional spectrum should behave similarly to each other, it raises a number of questions. First, there is no obvious connection between morphological uniformity and the ability to drop pronoun subjects, and second, it is not entirely clear what counts as a uniform paradigm.

A possible solution to the first problem may have something to do with the recovery of the identity of the null pronoun such as *pro*. A rich agreement system obviously helps us to recover the content of a null subject, as does the context in which the sentence is uttered: if we are talking about John and we say *went home*, it would be natural to assume that it was John who went home rather than someone else. For example in English many people keep diaries in which they omit the first person *I* (Haegeman, 1990):

(101) Got up. Had breakfast. Went to work.

simply because it is blindingly obvious who the subject of a diary is. If a language has no agreement morphology, this reliance on the context of situation would be the only way to recover the content of a null subject. It may be then that recovering the content of the null subject needs either inflection or discourse context, but that the latter is only available in the absence of distinct agreement inflection. Recovery of the content of a null subject in a language with agreement inflections can only happen if all possibilities are marked: on hearing the form *speak* we are not able to tell if the subject is first or second person, singular or plural.

Huang (1984) suggests that languages come in 'subject prominent' and 'topic prominent' types and that in the latter the notion of discourse topic plays a central role in the recovery of the identity of null elements. He points out that Chinese, which he claims is topic prominent, can have both null subject and null objects fairly freely and that unlike Italian the null object is not interpreted generically but is dependent on the topic:

(102) Zhangsan shuo Lisi bu renshi.
(Zhangsan say Lisi not know)
Zhangsan says that Lisi does not know him.

In this case, the missing object would be interpreted as identical to the topic, so if the conversation was about Bill, (102) would be interpreted as *Zhangsan says that Lisi doesn't know Bill*. The null object in this case, however, could not be interpreted as co-referential with the main clause subject, which is unexpected if it is simply a null pronoun, as an overt pronoun could refer to this subject. We might refer to this phenomenon as **topic-drop** rather than pro-drop. In Huang's theory, these sentences do not involve *pro*, however, but contain an empty topic operator which is moved to the front of the sentence; it is the reference of this operator that is fixed by the topic. As such the sentence is more like the English:

(103) Bill, John says Mary doesn't know.

with *Bill* unpronounced.

The second problem is more troubling. The verb paradigm we have been examining contains six entries for the present tense, being first, second and third person in singular and plural. But is it essential that there be uniformity across exactly these six? There are some languages which show gender distinctions as well as person and number; some have more than a singular and plural contrast, say the dual form found in Old English or three- or even four-way number distinctions. Is uniformity in these language defined with respect to the complete paradigm, or only with reference to the six forms mentioned so far? Perhaps not all slots in more complex paradigms have to be taken into consideration. For example Moroccan Arabic has nominally a 14-way agreement paradigm in the imperfect tense, including a dual form for all three persons and a masculine/feminine distinction in second and third persons singular and second and third persons plural, as shown in (104). From the start this does not appear to be uniform as gender distinctions are not uniformly made in all instances. Moreover not all of the 14 parts are morphologically distinct; items (f) first dual, (j) first plural and (l) second plural feminine are all the same form *tkəlm-na*:

(104)

a.	1st sing.	tkəlm-tu
b.	2nd sing. masc.	tkəlm-ta
c.	2nd sing. fem.	tkəlm-ti
d.	3rd sing. masc.	tkəlm
e.	3rd sing. fem.	tkəlm-at
f.	1st dual	tkəlm-na
g.	2nd dual	tkəlm-tuma
h.	3rd dual masc.	tkəlm-a
i.	3rd dual fem.	tkəlm-ata
j.	1st pl.	tkəlm-na
k.	2nd pl. masc.	tkəlm-tum
l.	2nd pl. fem.	tkəlm-na
m.	3rd pl. masc.	tkəlm-tum-na
n.	3rd pl. fem.	tkəlm-tum-at

By all accounts, this does not appear to be a uniform paradigm, but Moroccan Arabic is nevertheless a pro-drop language. However, the six greyed cases in (104), which ignore dual, show uniformity in that each is morphologically marked. While it appears that three person and two number distinctions are important for determining uniformity in a paradigm, it is not clear why this should be.

Another question concerns which paradigm counts when determining morphological uniformity. The examples so far all involve the present tense. Why should this take priority over other tenses? If the past tense paradigm is crucial, then English would come out as morphologically uniform:

(105) I/you/she/he/it/we/they looked.

(with the exception of the *was/were* contrast). Moreover, strictly speaking Hungarian is not morphologically uniform in its intransitive paradigm, as the third person singular form is the same as the base, for example *sétál* (walk). However, the paradigm for verbs such as *lát* (see) when they take a definite object is uniform in agreement:

(106)

	'walk' (intrans)	'see it' (def. object)
1st sing.	sétál-ok	lát-om
2nd sing.	sétál-sz	lát-od
3rd sing.	sétál	lát-ja
1st pl.	sétál-unk	lát-juk
2nd pl.	sétál-tok	lát-játok
3rd pl.	sétál-nak	lát-ják

Given that Hungarian is pro-drop, it must be the definite object paradigm that counts. But why should this be?

Finally, in some languages *pro* is allowed in some instances but not in others, as in modern Hebrew where pro-drop is possible in main clauses in past and future tenses with first and second person subjects, for example:

(107) ani axal-ti pro axal-ti
(I ate-1.sing.)

but not in present tense main clauses or main clauses with third person subjects (Borer, 1984):

(108) a. ani/at/ hi/ . . . oxelet **pro* oxelet
(I you she eats)
b. hem axlu **pro* axlu
(they ate)

As the third person is unmarked in all tenses and all persons are not distinguished in present tense, it seems that the possible appearance of *pro* in Hebrew is directly related to the richness of the particular inflection rather than to the uniformity of whole paradigms.

ALTERNATIVE APPROACHES TO PRO-DROP

Rizzi (1982): Rich agreement systems allow finite inflections to be proper governors and *pro* subjects are licensed by proper governors. Therefore pro-drop languages are those with rich agreement.

Rizzi (1986): What licenses *pro* is parameterized and languages select different possibilities (inflection, nouns, verbs, etc.). The content of *pro* must also be recoverable from its licensor, so rich agreement allows all null subjects to be recovered and poor agreement allows only expletive null subjects. Thematic licensors allow a generic interpretation of *pro* objects.

Huang (1984): Languages are either subject or topic prominent. *Subject prominent languages (such as Italian) can license pro with rich agreement, but* topic prominent languages (such as Chinese) can have empty topic operators associated with subject or object positions, which are dependent on discourse conditions.

Jaeggli and Safir (1989): Languages either have morphologically uniform (uniformly different or uniformly the same) or non-uniform inflections (some different, some the same). +uniform inflections license *pro*, –uniform inflections do not.

EXERCISE 3.4

1 Write down the present tense and past tense forms for a language you know other than English (if you know no other language find a grammar book for one). According to morphological uniformity, is this language likely to be pro-drop or non-pro-drop?

2 Here are the present tense forms for verbs in different languages. Assuming that the spelling reflects the pronunciation, do you think they are pro-drop or non-pro-drop languages?

Persian (a.k.a. Farsi)	**Dutch**	**Icelandic**	**Finnish**	**German**
'read'	'work'	'bite'	'ask'	'play'
1. mîkhânam	werk	bít	kysyn	spiele
2. mîkânîd	werkt	bítur	kysyt	spielst
3. mîkhânad	werkt	bítur	kysyy	spielt
4. mîkhânîm	werken	bítum	kysymme	spielen
5. mîkhânîd	werken	bíti	kysytte	spielen
6. mîkhânand	werken	bíta	kysyvät	spielen

We shall see in chapter 5 that pro-drop has provided a rich area for research and speculation about children's acquisition of their first language.

3.6 Further developments in X-bar Theory

This final section of the chapter introduces the developments in X-bar Theory during the 1980s that led to a still more general theory of structure.

3.6.1 *IP and CP*

Until now, the structural issues discussed have only concerned thematic elements such as nouns, verbs, adjectives and prepositions, i.e. the four lexical categories. But what about the other categories that have been mentioned only in passing, such as determiners, inflections and complementizers? These **functional categories** differ from the thematic categories so far discussed in that they play no direct role in assigning or receiving θ-roles. In fact, the semantic interpretation of such elements is often secondary to that of the thematic elements which express the basic content of the proposition. For example, consider the role played by the auxiliary verb *must* and the complementizer *that* in the following sentence:

(109) John said that he must leave.

The main content of the embedded clause asserts that someone identified by the pronoun *he* (perhaps John or perhaps someone else unnamed) is involved in an action expressed by the predicate *leave*. The modal auxiliary *must* adds the extra overtone that 'his leaving' is an obligation. It is not, however, clear that the complementizer *that* actually adds very much in the way of semantic content. To make this clearer, consider another sentence:

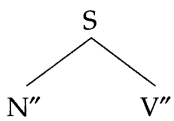
(110) John asked if he must leave.

The complementizer *that* in (109) demonstrates that the embedded sentence is declarative whereas the *if* in (110) shows that it is interrogative, i.e. based on a question. There is then a contrast between *that* and *if* as complementizers.

As thematic categories apparently play a more important semantic role in the meaning of the sentence, they are traditionally called **major** categories and the functional categories are often referred to as **minor** categories.

Until the 1980s it was assumed, following traditional grammar, that the syntactic importance of functional categories reflected their minor semantic importance and therefore was secondary to lexical categories. Hence X-bar Theory was about the thematic categories of nouns, verbs etc. and had little to say about functional categories such as complementizers. As no theory of the syntactic treatment of functional categories was developed, they were treated in a rather off-hand way, and were included in the positions where they seemed to fit best. At the same time, there was one major structure which seemed to be outside the X-bar system altogether, namely the sentence itself. The beginning of this chapter represented the clause as an 'S' dominating the subject NP and the VP:

(111)



But the fact that this S lacks a head means it clearly breaches the X-bar principle that every phrase must have a head.

In fact, structure (111) is already a simplification, as by the end of the 1970s a further position had been introduced to accommodate those elements that demonstrate finiteness or non-finiteness in the clause. These range from modal auxiliaries to the non-finite marker *to* and the finite inflection *-s*:

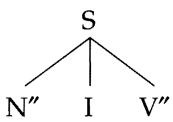
(112) a. John **will** leave.
b. I expect [John **to** leave].
c. John always leave**s**.

As every sentence needs one of these elements and they are in complementary distribution with each other, i.e. there is no possibility of:

(113) * John will leaves.

it was assumed that they form a single category making up a third obligatory part of the sentence. This category became known as **inflection** or INFL for short, later abbreviated simply to I:

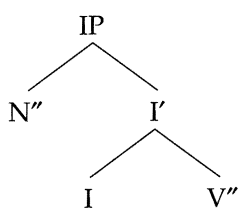
(114)



It is obvious how this I position accommodates modal auxiliaries and the infinitival *to*, as they appear in between the subject and the VP. It is more difficult to accommodate the inflections of tense and agreement, which attach to a verbal element. As this issue involves movement, it recurs in the next chapter, so we will ignore for now the fact that inflectional bound morphemes do not appear between the subject and the VP in the same place as other I elements.

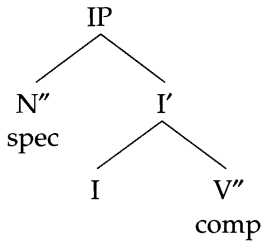
I is obviously a functional category rather than a lexical category. Hence it stands outside the X-bar conventions applying to lexical categories. But note that structure (114) includes a word-like element, the I, that is not the head of any phrase, and a phrase-like element, the S, which has no head. Given that clauses are categorized as finite or non-finite – exactly what the I element marks – it is easy to jump to the conclusion that I is in fact the head of the clause. At the start of the GB period the usual belief was as follows: 'Let us assume further that VP is a maximal projection and that the S-system [i.e. the clause] is not a projection of V but rather of INFL' (Chomsky, 1981a, p. 164). Only in the mid-1980s did it became regularly expressed in X-bar terms, using the diagram below:

(115)



This proposal claims not only that the I element is the head of the clause, and hence that the clause is the maximal projection of the inflection, an IP, but also that the VP is the complement of the inflection and the subject its specifier:

(116)



This offers a straightforward account of the word order of the English sentence in terms of X-bar syntax, as the complement follows the head while the specifier precedes it. There is also structural evidence that the inflection and the VP form a single constituent in the sentence as they may coordinate with other I + VP sequences:

(117) He *may attend the conference* but *won't present a paper.*

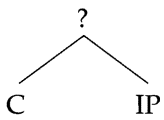
Furthermore, only VPs can follow inflections, indicating a restrictive relationship between them, and supporting the claim that the VP is the complement of the inflection.

As soon as one functional element is seen to take part in the X-bar system, the pressure is on to assume that they *all* do, leading to a general theory of structure in which all elements conform to X-bar Theory, whether functional or lexical. Around the same time that the inflection was accommodated into the X-bar framework, similar ideas were therefore proposed for the complementizer. As pointed out earlier, complementizers are elements like *that* which introduce embedded clauses and they usually carry features distinguishing finite and non-finite as well as declarative and interrogative. For example, the complementizer *that* introduces finite declarative clauses while *for* introduces non-finite declaratives:

(118) a. I think *that* he may know.
b. I was anxious *for* him to know.

As we have seen in (110), the complementizer *if* introduces finite interrogatives. During the 1970s, it had been established that the complementizer (C) forms a constituent with the clause that it introduces; however, C was not included with the subject, the inflection and the VP as part of the basic clause structure. If the basic clause is in fact an IP, this indicates the following structure:

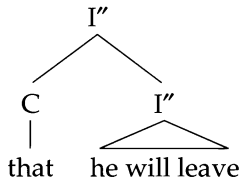
(119)



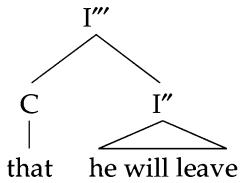
There are three options for the identity of the constituent labelled with the question mark in tree (119): it might not be a headed constituent at all, in which case

it does not fall within the remit of the X-bar system; it might be headed by the IP; or it might be headed by the complementizer. The second of these options is the least feasible as it would either place the complementizer in an adjunction position, adjoined to the IP, or would require extending the IP to a third projection level:

(120) a.



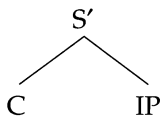
b.



As the complementizer is not recursive, the adjunction proposal in structure (120a) of having one I″ within another I″ is improbable, while the proposal in structure (120b) that clauses extend to a triple bar level I‴ undermines the claim that the inflection conforms to the same theory of phrase structure developed for the thematic elements: nouns and verbs, etc., project only two levels according to standard X-bar Theory.

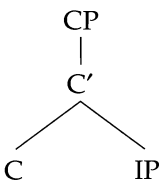
What remains are the proposals that the constituent containing the complementizer is either headed or not headed. If it isn't headed, then we recreate a similar position needed for the analysis of the 'S' node: on the one hand there is a word-like category, the complementizer, which heads no phrase, on the other a phrase-like element, the constituent containing the complementizer and the IP – let's call it S′ (as it was called during the 1970s) – which has no head:

(121)



Once more this configuration strongly suggests that the complementizer should be taken as the head of the constituent, as is also suggested by the fact that the complementizer plays a role in determining the declarative or interrogative status, what may be referred to at the **force**, of the entire clause. Thus, we assume the following structure:

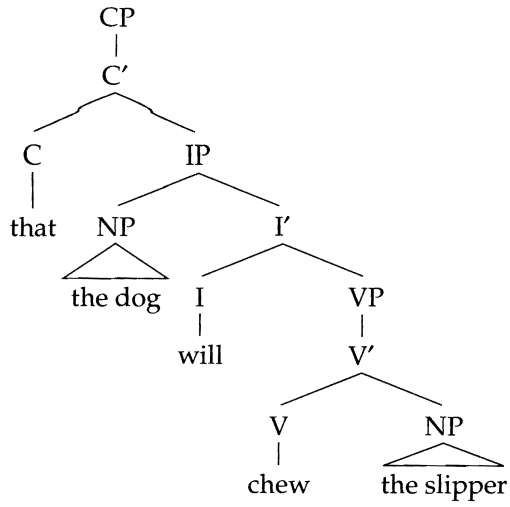
(122)



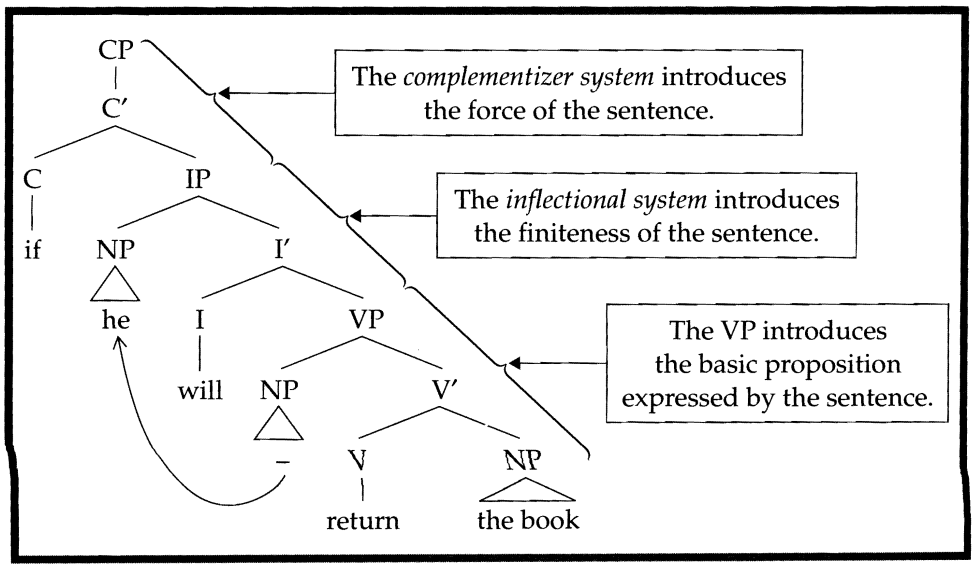
For the time being, we will skip over what counts as the specifier of the CP, returning to it in the next chapter. Note that structure (122) claims that the IP functions as the complement of the complementizer; this claim is supported by the observation that only an IP can follow a complementizer, indicating a restrictive relationship between them.

The full structure of the clause that we end up with is then as below:

(123)



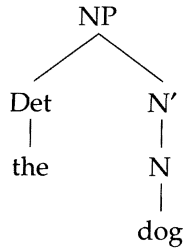
The clause is now structured into three hierarchical parts, all of which conform to usual X-bar Theory. At the bottom is the VP supplying the basic thematic elements which make up the proposition. Recall that under the VP-Internal Subject Hypothesis, the underlying position of the subject is in the specifier of the VP. So the VP contains the verbal predicate and all its arguments at the level of D-structure, though the subject moves out of the VP at S-structure. Next above the VP is the inflectional IP system which provides the distinctions of finiteness through the introduction of an inflectional head I. Its specifier is the surface position of the subject, as represented in (123). Finally at the top comes the complementizer CP system which introduces the force of the clause: i.e. whether it is declarative or interrogative.



3.6.2 *The DP Hypothesis*

Another functional element to be reanalysed within a more general X-bar Theory was the determiner. The standard assumption within phrase structure theory was always that determiners such as *the* and *a* sit in the specifier of the NP:

(124)

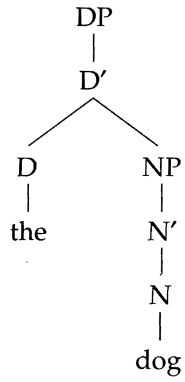


Once again we encounter a word-like element that is not the head of a phrase and therefore seems not to take part in X-bar Theory. There is also something strange about it sitting in the specifier of the NP. This position is the one which the possessor of the NP is thought to occupy, e.g. *John* in *John's dog;* the fact that determiners and possessors are in complementary distribution in English supports the assumption that they occupy the same position. However, the possessor is a phrase while the determiner is a word, and in other structures it seems that word positions and phrase positions are strictly demarcated. Indeed, in other languages the determiner and the possessor are not in complementary distribution. For example, in Hungarian it is quite common to find both together in the same NP:

(125) a Józsi kutyája
(the Józsi dog-3rd sing.)
Józsi's dog

But, if the determiner does not sit in the specifier of the NP, where does it sit? Moreover, if the determiner conforms to X-bar Theory, there must be a phrase that it heads, but what is that phrase? These puzzles were solved by Abney (1987) with the **DP Hypothesis**, which claims that the determiner is the head of the nominal phrase, not the noun. From this perspective, the noun heads a phrase which acts as the complement of the determiner:

(126)



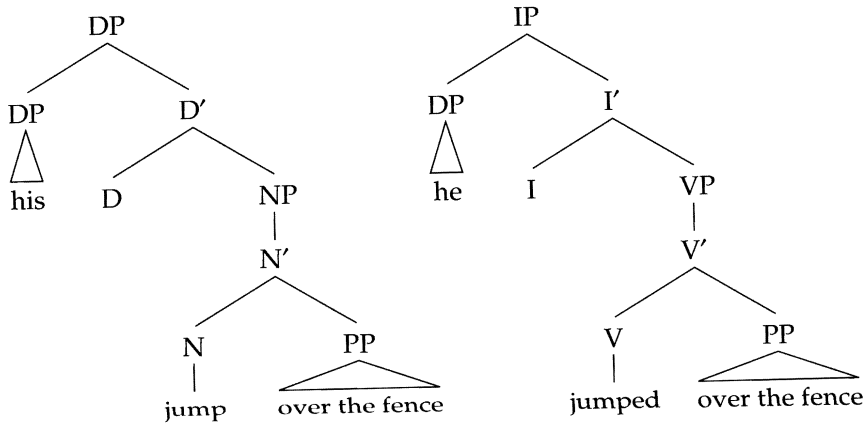
Again, there are good reasons to consider the determiner as a head of the nominal phrase. For example, it is often the determiner and not the noun which contributes the property of definiteness or indefiniteness to the whole phrase:

(127) a. the dog
b. a dog

The phrases in (127a) and (127b) are definite and indefinite respectively, but the noun is identical in both cases. The distinction between the two lies with the determiners and therefore it is the determiner which projects its properties of definiteness to the phrase rather than the noun.

A number of advantages follow from the DP analysis. One is that the structure of the nominal phrase exactly mirrors that of the IP:

(128)



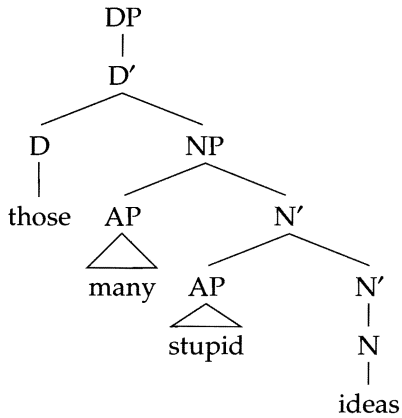
We commented above on the similarity between the subject of the clause and the possessor of the nominal phrase, which was one of the motivations for the development of X-bar Theory in the first instance. Under the DP Hypothesis, the structural parallelism between the two is exact: both sit in the specifier of a functional element which selects as a complement the phrase headed by the thematic element to which the subject and possessor are related.

This analysis also increases the number of structural positions within the nominal phrase, another advantage: not only do we separate the possessor from the determiner, but we also introduce a second specifier position – that of the noun. Abney (1987) argued that this position is needed to accommodate what are traditionally called the *post-determiners* – determiner-like elements which typically come after standard determiners:

(129) those *many/few/several* ideas

These elements cannot be attached within the traditional NP analysis, as the only place for them would be adjoined to the N′. But they are not recursive and so are not well analysed as adjuncts. Furthermore, they always precede adjectival modifiers, which are adjoined to N′. The NP specifier position within the DP analysis provides us with an ideal place to site post-determiners:

(130)



While the DP Hypothesis accounts for those languages in which possessors and determiners are not in complementary distribution, it does of course raise the problem of accounting for this pattern in languages where they are. As the possessor and the determiner do not occupy the same structural position, there must be some other reason why they cannot co-occur. Abney suggests that an abstract determiner obligatorily accompanies the possessor, a possessive determiner (*pos*), and that this is the element that is in complementary distribution with other determiners:

(131) a. his *pos* dog
b. the dog
c. * his *pos*/the dog

A final advantage of the DP Hypothesis is that it allows a principled account of the complementary distribution between determiners and pronouns. If, as is usually assumed, the noun is the head of the nominal phrase, one might have thought that pronouns replace the noun when they pronominalize an NP. But in this case, we might expect that pronouns could appear with an accompanying determiner, like other nouns. This is not so:

(132) * the him

However, this is not the same kind of restriction as with certain classes of nouns, such as proper nouns for example, which usually have no determiners, but may appear with them under special circumstances:

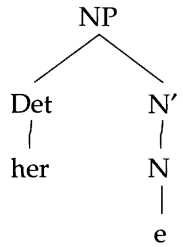
(133) a. * The Mary left.
b. She's not the Mary that I used to know.
c. There's a Mary at the door.
d. I know three Marys.

Pronouns, however, cannot occur with determiners under any circumstance:

(134) a. * The she left.
b. * She's not the her that I used to know.
c. * There's a her at the door.
d. * I know three hers.

One possible account of this pattern would be to assume that pronouns replace not the noun, but the determiner. But, under the NP analysis, this would involve proposing an abstract noun head, which has no motivation at all given that the entire content of the NP is provided by the pronoun in the specifier position:

(135)

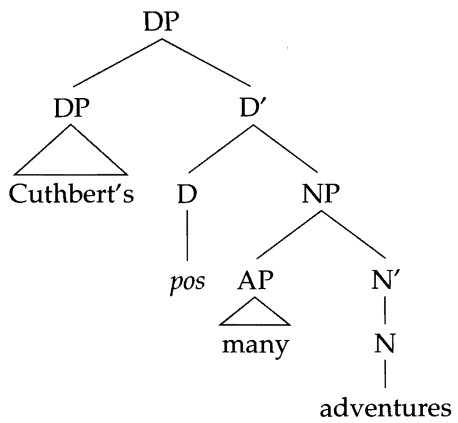


The DP Hypothesis, on the other hand, maintains the assumption that pronouns replace determiners without the need to postulate an abstract noun. Given that the determiner is the head of the nominal phrase, it is the only element that is required in the DP. In other words, we analyse pronouns as 'intransitive' determiners:

(136)

THE DP HYPOTHESIS

According to the DP Hypothesis, the determiner, not the noun, is the head of the nominal phrase. The possessor is in the specifier of the DP and the phrase headed by the noun is the complement of the determiner. The specifier of the NP is the position of the post-determiner:



EXERCISE 3.5

Provide possible trees for the following DPs in Hungarian:

Csaba háza (Csaba house-3rd sing.) Csaba's house	Attilának a tehene (Attila-dative the cow-3rd sing. Attila's cow
Csilla minden síbotja (Csilla every ski-stick-3rd sing. all of Csilla's ski sticks	a Józsi sok hibája (the Józsi many mistake-3rd sing. Józsi's many mistakes

Which of these is problematic for the analysis discussed above? Could this problem be solved if we supposed that the Hungarian DP had a further level of functional projection similar to the CP of the clause?

3.6.3 *The Split INFL Hypothesis*

A further important development within X-bar Theory again concerns the inflectional elements. In English the inflections of the verb indicate tense and subject agreement features, though, as we have pointed out, distinctions for subject agreement are particularly poor in English. Importantly the verb has just one morpheme attached to it which expresses either past or present tense with third person singular agreement:

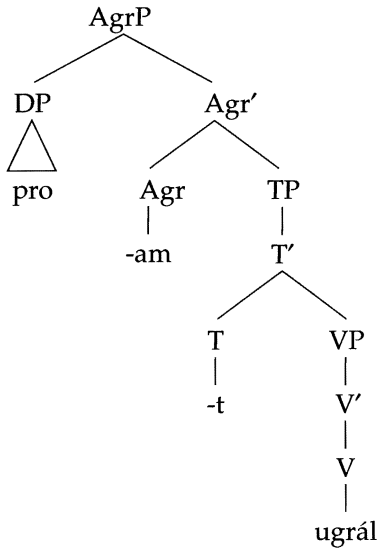
(137) a. He/I/you/ . . . jump**ed**.
b. He jump**s**.

This corresponds to the claim that there is just one inflection node in the clause. However, in languages other than English tense and agreement are represented by independent morphemes. Consider the following Hungarian example where not only is past tense represented by '-t' but also the first person singular agreement is shown as '-am':

(138) Ugrál-t-am.
(jump-past-1.sing.)
I jumped.

If there is only one inflection node in a sentence, where do the separate inflectional morphemes fit? One possibility is that they occupy their own head position and project their own phrase, i.e. AgrP (Agreement Phrase) and Tense Phrase (TP):

(139)



Pollock (1989) argued that a structure similar to this is in fact universal, even for languages without separate tense and agreement morphemes. His evidence comes from French. Assuming inflections are generated as separate elements from the verb and that these come together as the result of some syntactic process, there are two possible surface positions in which the verb might occur: in the verb position, with the inflection adjoined to the verb, or in the inflection position, with the verb adjoined to the inflection. This seems to be a parametric difference between languages. For example, the following sentences suggest that English verbs remain inside the VP while French verbs move to the inflection position:

(140) a. John [$_{VP}$ often [$_{VP}$ kisses Mary]].
b. Jean embrasse [$_{VP}$ souvent [$_{VP}$ – Marie]].

Assuming that the adverb is adjoined to the VP, in the English example the verb *kisses* sits between the adverb *often* and the object *Mary*, showing that it remains in the verb position. In French, however, the verb *embrasse* precedes the adverb *souvent* indicating that it is no longer within the VP, but in the position posited for the inflection, namely between the VP and the subject. Essentially the same facts concerning the verb seem to hold in negative sentences, though things are slightly more complicated in both English and French: English verbs remain inside the VP and French verbs sit in the inflection position:

(141) a. John did not [$_{VP}$ kiss Mary].
b. Jean n'embrasse pas [$_{VP}$ – Marie].

In the English example, the verb *kiss* sits between the negative element *not* and the object *Mary*, indicating a VP internal position. However, the dummy auxiliary *do* is inserted into the inflection position. We will discuss this phenomenon in the next chapter; we will overlook it for the moment as it has little bearing here. In

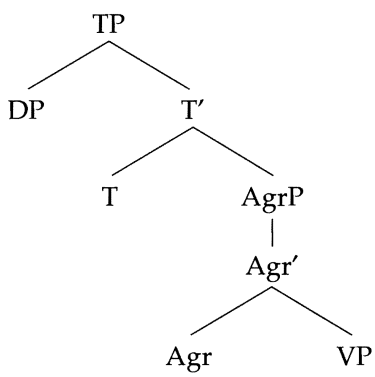
the French example, the verb *embrasse* is to the left of the negative element *pas*, which we might suppose is the equivalent of the English *not*; again the indications are that the verb is in the inflection position. We also have a second negative morpheme, the *n'*, which acts as a clitic stuck to the front of the verb (not to be expanded on here). The reason for introducing the negative sentences is to demonstrate that there is another possible position for the verb to occupy between the verb and inflection positions. This can be seen in French infinitival clauses in which the verb never sits in the inflection position, but may occupy a position to the left of a VP adjoined adverb:

(142) Ne pas embrasser [$_{VP}$ souvent – Marie] c'est triste.
(not to-kiss often Mary is sad)
To not often kiss Mary is sad.

In this example, the verb *embrasser* is to the right of the negative *ne pas*, showing that it is not in the inflection position, as it is in the finite clause in (141b). However, it *is* to the left of the adverb, showing that it is not in the VP either. But if the verb *embrasser* is neither in I nor in V, where is it? Pollock argued for the existence of another head position between the I and V positions, which he took to be the agreement head position, even though French does not have separate tense and agreement morphemes. Thus, he argued that something similar to (139) is applicable to French and, indeed, universally.

The structure that Pollock actually argued for is slightly different to (139), however, in that he positioned the AgrP below the TP:

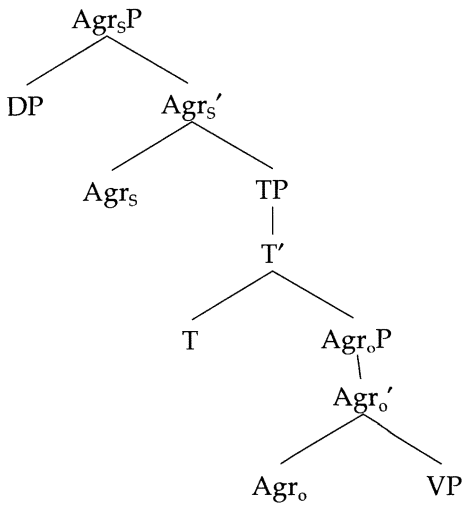
(143)



Pollock's argument was that the difference between French finite and infinitival clauses on which he based his analysis is a difference in the tense head T, not the agreement. In the finite clause, the verb is in a position that it cannot occupy in the infinitive clause and this suggests that French finite tense and infinitive tense have different properties. However, as the verb is in the highest position in the finite clause, but not in the highest position in the infinitival, this suggests that tense is the highest inflectional head. Other linguists, on the other hand, argued equally strongly for the structure in (139). For example, the order of the Hungarian morphemes suggests that the verb attaches first to the tense and then to the agreement, arguing that the tense node is nearer to the verb than the

agreement. A solution was proposed by Chomsky (1995a), who claimed that both parties were right: there is an agreement head above tense and one below tense. The highest agreement element is associated with agreement with the subject and the lowest one is associated with agreement with the object, which some languages show morphologically. Thus, Chomsky's solution is as follows:

(144)



Towards the end of the 1980s the idea that there was a rich system of functional heads built on top of the VP gave rise to much research. The idea became known as the **Articulated INFL Hypothesis**. Currently the status of such structures is arguable. Some researchers still assume such a complex system of functional projections, while others have retreated to a more minimal set. Chomsky (1995a) himself has argued that agreement, both subject and object, should be viewed as features rather than structural nodes, and has essentially reverted to a version of the CP/IP analysis of the clause. 'As matters stand here, is seems reasonable to conjecture that Agr does not exist and that [agreement]-features of a predicate P . . . are added optionally as P is selected from the lexicon. Note that this carries us back to something like the analysis that was conventional before Pollock's (1989) highly productive split-I theory' (Chomsky, 1995b, p. 377).

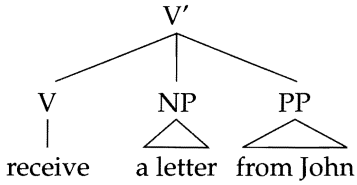
ARTICULATED INFL

Towards the end of the 1980s, what had been thought of as a single structural node for the inflectional elements was reanalysed as being made up of a number of separate inflectional heads, TP, AgrP, etc., each of which projects its own phrase. There was argument over whether the agreement phrase was above or below the tense phrase, until Chomsky (1995a) suggested that there are two agreement phrases, one above and one below the tense phrase.

3.6.4 The VP Shell Hypothesis

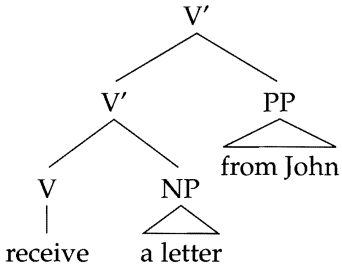
Finally we turn to developments concerning the structure of the VP. All of the structures drawn so far have involved either a transitive or an intransitive verb: i.e. verbs with either zero or one complement. However, verbs may take more than one complement – so how can two or more complements be accommodated within an X-bar structure? Two obvious solutions come readily to mind. First, if all complements are sisters to the head, multiple complements should all be at the same structural level, under the first projection of the head:

(145)



However, this structure involves a node with three branches, something not needed previously. The fact that binary branching trees are sufficient for most structures reflects a restriction on possible phrase structures known as the **Binary Branching Condition**, which states that structures are *at most* binary branching. This would then rule out a structure such as (145), because it has three branches. An alternative would be to reject the assumption that all complements are sisters to the head and allow some to be generated in what we have been calling the adjunct position:

(146)



This structure has a number of advantages over (145) apart from its conformity to the Binary Branching Condition. Most importantly, if we give up the definition of complements as sisters of the head and adjuncts as those elements which sit in adjoined positions, we reduce a redundancy in the previous accounts, as the notions 'complement' and 'adjunct' can also be defined in terms of their semantic relations to the head: complements are selected by the head and therefore appear in the head's subcategorization frame, whereas adjuncts are non-selected modifiers and do not appear in the subcategorization frame. The Sisterhood Condition on thematic role assignment ensures that complements are inserted into a structure closer to the head than adjuncts, though we must assume that thematic roles that cannot be assigned directly to the sister of the head will be passed on to its projection to be assigned to its sister. This does not add any extra grammatical mechanisms as we already assume the same thing for the external θ-role.

Despite the advantages of (146) over (145), there are reasons to believe that even this structure is not correct. As pointed out by Larson (1988), several phenomena seem to show that the first complement is structurally higher than the second, whereas (146) has the second argument in the structurally higher position. Here we will demonstrate just two of these phenomena. The first involves the referential properties of reflexive pronouns, the details of which will be provided in the next chapter. For now, all that is important is to note that a reflexive pronoun can be referentially dependent on an argument which is higher than itself, but not lower. So a reflexive object can refer to a subject, but not vice versa. In the following examples, we again use co-indexation to represent co-reference:

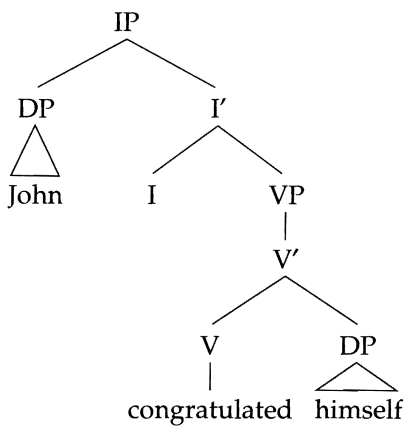
(147) a. John$_i$ congratulated himself$_i$.
b. * Himself$_i$ congratulated John$_i$.

The relationship involved in this phenomenon is known as **c-command**, which can be defined in the following way:

(148) **c-command**
An element c-commands its sister and everything dominated by its sister.

To clarify, consider the following tree diagram:

(149)



The sister of the subject *John* in this structure is the I′. As the object *himself* is included in the I′, the subject c-commands the object. The sister of the object is the V and therefore the object c-commands the V. The subject is obviously not included in the V and so the object does not c-command the subject.

Thus we can say that a reflexive pronoun can only refer to a c-commanding element. When a verb has two objects, the second object can be referentially dependent on the first, but not vice versa:

(150) a. John showed Bill$_i$ himself$_i$ (in the mirror).
b. * John showed himself$_i$ Bill$_i$ (in the mirror).

This suggests that the same structural relationship holds between the two objects as between a subject and an object. As the subject c-commands the object, the first object therefore must c-command the second. Going back to structure (146), this makes exactly the opposite claim: here the second complement c-commands the first. We might therefore conclude that this structure cannot be correct.

A second observation which shows that the first object of a double object predicate c-commands the second concerns negative polarity items, such as *anyone*. Such elements can appear if they are c-commanded by a negative element. Thus *anyone* can function as the object of a sentence with a negative subject, but *anyone* cannot occur in subject position with a negative object:

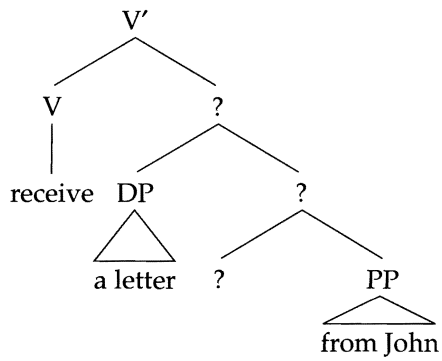
(151) a. No one likes anyone.
b. * Anyone likes no one.

In a double object construction, a negative polarity item can appear in the second object position with a negative element in the first object position, but not vice versa, demonstrating that the first object c-commands the second:

(152) a. I told no one anything.
b. * I told anyone nothing.

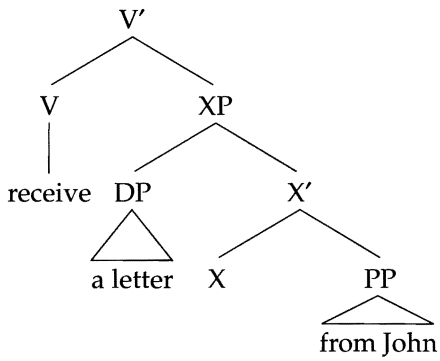
However, for the first object to c-command the second would require a structure something like the following:

(153)



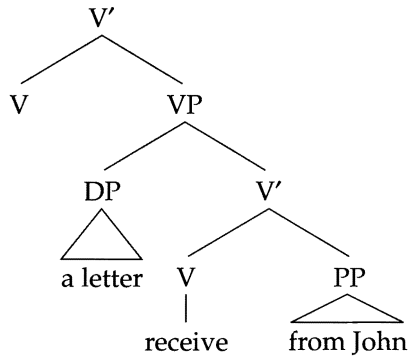
The part of this structure with the labels '?' looks suspiciously like a phrase with the DP in its specifier position and the PP in its complement position:

(154)



The question is: what is the head of this phrase? Larson (1988) proposed that the head is the verb and that this gets into its preceding position via a movement. Thus, underlyingly the phrase looks like the following:

(155)



This analysis additionally proposes an empty shell of a VP generated on top of the contentful VP to provide a landing position for the verb to move to: hence it is called the **VP Shell Hypothesis**.

Others have suggested that the head of the VP shell is not always just an empty position for the verb to move to, but may contain an abstract verbal element which the overt verb attaches to via movement. Thus consider the following observations:

(156) a. The ship sank.
b. They sank the ship.
c. They made the ship sink.

Verbs such as *sink* are sometimes called **ergative** verbs and appear to have both an intransitive and a transitive use as in (156a) and (156b). The difference is that the transitive use involves causation – something made the ship sink – as shown by the fact that (156b) and (156c) mean similar things. The subject here is interpreted as the one that causes the sinking to happen. In the intransitive example, however, the subject is the thing that undergoes the sinking. The word order facts and the near synonymy of (156b) and (156c) can be accounted for under the following assumptions. First, the basic position for the element that undergoes the sinking is the specifier of the verb, thus accounting neatly for (156a) and (156c):

(157) a. $[_{VP}$ The ship $[_{V'}$ sank]].
b. $[_{VP}$ They made $[_{VP}$ the ship $[_{V'}$ sink]]].

(157a) shows a partial structure, ignoring the functional structure that would be built above this VP and the subsequent movement of the subject to the specifier of the IP. The causative verb *make* is assumed to take a VP complement identical

in all relevant respects to the VP in (157a). Sentence (156b) raises a number of questions: why is this structure interpreted as a causative when there appears to be no causative verb? And why is the theme argument *the ship* after the verb in (156b) but before it in (156a)? One answer to these questions assumes there is a causative verb in (156b) which is phonologically empty and the overt verb moves to attach to it:

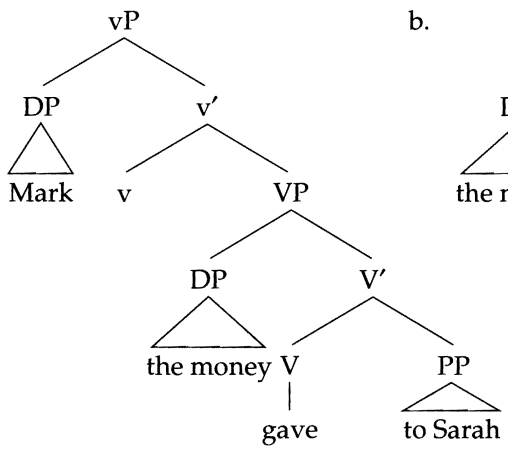
(158) a. [$_{VP}$ They e [$_{VP}$ the ship [$_{V'}$ sank]]].

b. [$_{VP}$ They sank-e [$_{VP}$ the ship [$_{V'}$ –]]].

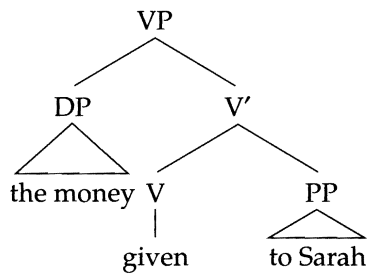
Note that this is exactly the same kind of movement found in the VP-shell hypothesis with a verb moving from a lower verbal position to a higher one. Here, however, the upper verbal position is not just an empty position for the verb to move to, but contains a meaningful verb of its own.

More recently, this line has been pursued further, suggesting that the verb of the VP shell is what is responsible for assigning the external θ-role in general, particularly the agent. This verb, known as a *light verb* (typically represented as a lower case 'v'), is present only when an agent role is assigned, and hence one possible analysis of the active/passive distinction might be in terms of the presence or absence of the light verb:

(159) a.



b.



The structure in (159a) represents the core proposition of a sentence that would be realized as:

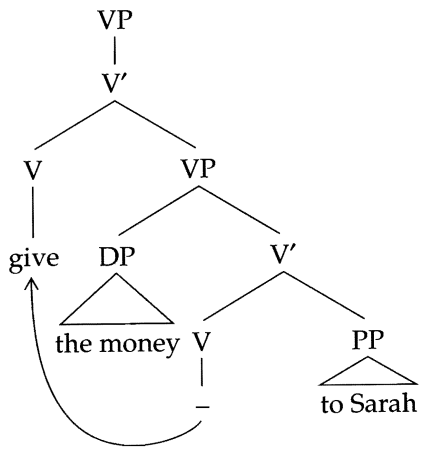
(160) Mark gave the money to Sarah.

whereas (159b) represents the proposition of the passive sentence:

(161) The money was given to Sarah.

VP SHELLS

In order to maintain the Binary Branching Condition in the case of verbs with multiple complements, and to capture the observations that show that the first complement c-commands (is structurally higher than) the second, Larson (1988) proposed that the VP structure should be articulated into two parts: an upper VP shell, empty of lexical elements, and a contentful VP containing the verb and its complements in specifier and complement positions. The verb is then assumed to move to the head position of the VP shell:



We will conclude this section, and this chapter, by briefly mentioning that just as the IP and VP have undergone an 'articulation' process, analysing them into a series of functional projections headed by more specific heads, such as tense, agreement and causative verbs, the CP has also been subject to a similar analysis. Rizzi (1997) has argued that what has been referred to as a single projection headed by the complementizer is better seen as a number of separate projections of heads which introduce notions such as topic and focus. The suggestion remains contentious, especially in the light of the recent move towards simplification of the clausal architecture, almost returning to a simple IP analysis. However, some linguists are still following the line of research started in the late 1980s, which has become known as the 'functional explosion'. As Chomsky himself has advocated a more minimal approach, we will not pursue this further.

3.7 Summary

This chapter has concentrated on issues of structure that developed throughout the 1980s within GB Theory. Many of these serve as precursors to the Minimalist

Program, to be discussed in chapters 7 and 8. Most of the discussion so far concerns the development of X-bar Theory, which became increasingly general throughout the period of GB Theory, starting off as a theory of the structure of phrases headed by lexical elements, N, V, A and P, and ending up as a completely general theory of all syntactic structure. The most important development in this respect was extending X-bar Theory to cover functional elements, such as complementizers, inflections and determiners. From these the interest in functional elements and their role in syntax increased and later developments such as the Articulated INFL Hypothesis followed. There were also developments in thinking about the structure of the VP, in particular the VP-Internal Subject Hypothesis and the VP Shell Hypothesis, both of which contribute to a view of the VP as a hierarchical structure with all but the lowest argument being included as specifiers of VP shells built one on top of another.

Other modules of the theory with special relevance for D-structure, i.e. Theta Theory and Control Theory, have also been discussed. Theta Theory deals mainly with the assignment of thematic roles from predicates to arguments and contains two basic principles: the Theta Criterion and the Sisterhood Condition, which play a role in regulating how and where θ-roles can be assigned and therefore determining which arguments appear where. Control Theory introduced the notion of phonologically empty arguments, which in many ways behave like pronouns in terms of their distribution and referential properties. Control Theory itself is concerned with the referential properties of PRO, an empty category which always sits in ungoverned positions, according to the PRO theorem (79). We contrasted this with another empty pronoun, *pro*, which sits in governed positions in certain languages.

Due to the highly interactive nature of the modules of GB, it is virtually impossible to discuss any single module without reference to the others. This chapter has made numerous references to modules of the grammar which will be more fully discussed in the next chapter. The modules that are relevant for S-structure (and beyond) come into their own as conditions on what movements can or must take place.

Discussion topics

1 To what extent do you think deconstructing the theory into a proliferation of modules simplifies or obfuscates?
2 Does the distinction between deep and surface structure strike you as obvious and familiar, or do you find it novel and startling?
3 Do you feel any of the claims about English reflected in the rewrite rules are discoveries about English or are formalizations of already known grammatical rules?
4 How would you now define the subject of the sentence? How does this compare to your definition before reading this chapter?
5 To what extent do you think that the fact that the generation of the sentence is now driven by the lexicon diminishes the unique status of syntax?
6 Recently Hauser et al. (2002) seem to be suggesting that the only unique aspect of the 'narrow' language faculty is recursion. Can this really be so important?

7 The most widely used artificial language in the world, Klingon, allegedly with one native-speaker baby (http://www.tvwiki.tv/wiki/D%27Armond_Speers), has no tense forms yet appears to be non-pro-drop, as in *naDev tlhInganpu' tu'lu* '(There) are Klingons here': http://stp.ling.uu.se/~zrajm/nerd/klingon/TKDAddenda/kliadd4.html#4. Is this because its devisers were chiefly English speakers or because it is actually an invented language, i.e. not subject to UG?

4 Movement in Government/ Binding Theory

The previous chapter introduced the idea that the structure of an expression in any language can be described at two levels: D-structure and S-structure. These levels are linked by transformational rules, which mostly involve moving elements from one place to another – movement, alias dislocation. The levels are also the places at which various principles, assembled into modules, apply. The principles that apply at D-structure govern the basic organization of the structure, as we have seen in the last chapter, while the principles that apply at S-structure regulate movement. For example, an S-structure principle may require an element to be in a certain position which it does not occupy at D-structure. So, for the sentence to be grammatical, the element will have to vacate its D-structure position and move to the position it is required to be in at S-structure. There are also occasions where S-structure principles prevent certain movements from happening, thus explaining why some movements are possible, others not. An interesting aspect of the Government/Binding (GB) grammar is that many principles are not motivated solely by considerations of movement, but also have roles to play in other phenomena. Before these can be discussed in detail, however, we should first look more thoroughly at how the notion of movement has developed in Chomskyan linguistics over the years.

4.1 An overview of movement

4.1.1 A-movements

As discussed in the previous chapter, the notion of a transformation started off as a powerful device that could alter structures in unconstrained ways. However, it was clear from the start that, to attain the goal of explanatory adequacy, this flexibility could not be maintained, as no real theory of language acquisition would be possible without restrictions on the mechanisms that grammars could potentially utilize, i.e. children would simply be unable to learn language if any

structure at all was possible. Building in restrictions meant that the transformational rules themselves could be simplified and generalized.

By the late 1970s three types of movement transformations were recognized, which played a role in the generation of numerous structures (Chomsky, 1977). One of these types involves the movement of a Determiner Phrase into a subject position. Several constructions involve this kind of movement, the most obvious being the passive. In a typical passive, a DP originates in the object position and then moves to the subject which is vacant:

(1) a. – was amended the contract.

b. The contract was amended –.

This analysis, which dates back to the first transformational analyses of the 1950s, implicitly assumes a principle made explicit by Baker (1988), called the **Uniform Theta-Role Assignment Hypothesis**, or UTAH for short: arguments which bear the same θ-role are generated in the same D-structure position. Thus, if arguments which are interpreted similarly occupy different surface positions, the difference must be attributed to movement. This is the case for instance with the comparison of passive and active sentences, where the object of the active and the subject of the passive bear the same θ-role:

(2) a. The lawyer amended the contract.
b. The contract was amended by the lawyer.

A similar movement can be seen in the so-called middle construction:

(3) These potatoes mash well.

This construction is restricted to a smaller set of transitive verbs, and has other restrictions which need not concern us here. The point is that an element which is interpreted as the object of the Verb (*these potatoes*) ends up in the subject position. We might propose therefore that this movement is similar to the passive, with the object moving to the vacant subject position:

(4) a. – mash these potatoes well.

b. These potatoes mash – well.

A further example of this type of movement involves the so-called *unaccusative verbs* such as *arrive*. While these appear to be intransitive at the surface, they have some properties of transitive verbs. In English the most obvious difference is that unaccusative verbs can appear in the '*there* construction' but intransitive verbs cannot:

(5) a. There arrived a party from Cricklewood.
b. * There telephoned a party from Cricklewood.

In this construction, the subject is taken by the pleonastic element *there*. The element that would normally be interpreted as the subject (*a party from Cricklewood*) appears after the verb, in what seems to be an object position. Another difference is that intransitive verbs can take a special object which in some sense duplicates the meaning of the verb, known as a **cognate object**:

(6) a. He died a terrible death.
b. He smiled a rueful smile.
c. He danced a merry dance.

Unaccusative verbs, however, do not take cognate objects, a property they share with transitive verbs:

(7) a. * He arrived an arrival.
b. * They destroyed a destruction.

Presumably the reason why transitive verbs cannot take cognate objects is because these objects cannot really bear θ-roles and, according to the θ-Criterion seen in chapter 3, a verb which has a θ-role to assign to an object must do so. Unless the same is true for unaccusative verbs, i.e. that they have a θ-role to assign to an object, it is unclear why they should not be able to take a cognate object.

Data from other languages also show that unaccusative verbs have certain things in common with transitive verbs. For example, in Italian, the pronoun *ne* can be cliticized in front of the verb if it is associated with the object but not with a postposed subject (recall from the previous chapter that Italian is a pro-drop language, and one property of some pro-drop languages is that the subject can follow the VP):

(8) a. Luigi ne ha insultati molti.
(Luigi of-them has insulted many)
Luigi has insulted many of them.
b. * Ne telefonano molti.
(of-them telephone many)
Many of them telephone.

However, we can get *ne*-cliticization from an apparently postposed subject of an unaccusative verb:

(9) Ne arrivano molti.
(of-them arrive many)
Many of them arrive.

Thus in (9) there is no postposed subject, for this element is really an object.

All this adds up to the conclusion that unaccusative verbs are really transitive verbs, selecting for a single complement. However, this complement may move to the subject position, just like the objects of passive and middle verbs:

(10) a. – arrived a party of linguists.

b. A party of linguists arrived –.

The movement of a Determiner Phrase to a subject position is also found with **raising verbs**. The difference is that the Determiner Phrase moves from another subject position rather than from an object position:

(11) a. – seems [John to have lost his passport].

b. John seems [– to have lost his passport].

At first sight, an S-structure such as (11b) looks identical to those introduced in the previous chapter which involve a controlled empty pronoun, PRO. Compare the following:

(12) a. John seems to have lost his passport.
b. John expects to have lost his passport.

However, the many differences between these two constructions lead us to believe that (12a) involves a movement while (12b) does not. First of all the subject in (12a) is not thematically related to the verb *seem*: it is not *John* that seems, so to speak, but 'that John has lost his passport', as seen clearly in the following sentence, which, while synonymous with (12a), does not involve the movement of the subject:

(13) It seems that John has lost his passport.

Here the subject of *seems* is an expletive *it*; the DP *John* is semantically related to *lost* and so must be generated in the *lost* clause at D-structure in both (12a) and (13) according to the Sisterhood Condition of Theta Theory (chapter 3, section 3.4). Turning to (12b), in this case the S-structure subject is thematically related to the Verb of the main clause: it *is* John that is doing the expecting. This DP must then have been generated in the main clause at D-structure and so did not move there from the embedded clause. Indeed there is no clause similar to (13) in this case:

(14) * It expects that John has lost his passport.

This clause is only grammatical if the subject is interpreted as referential, i.e. linked to an actual thing, and not as an expletive, showing that the verb *expect* is lexically different from the verb *seem*: while *expect* has a thematic subject of its own, *seem* does not. Their lexical entries can then be represented as follows:

(15) a. seem: [CP], <theme>
b. expect: [CP], <experiencer, theme>

While both verbs subcategorize for a clausal complement, *seem* has no other arguments while *expect* has an experiencer subject. Verbs like *seem*, which have

no subject of their own but allow the subject of their clausal complement to move to their own subject position, are known as **raising verbs**.

So far we have seen that passive, middle, unaccusative and raising verbs all allow DPs to move to their subject position and that the processes involved are very similar. In fact, if we accept the VP-Internal Subject Hypothesis, this type of movement seems to happen in virtually every clause. The assumption is that the subject originates in the specifier of VP at D-structure and then moves to the 'canonical subject position', i.e. the specifier of IP:

(16) a. – will [the princess kiss the frog].

b. The princess will [– kiss the frog].

This involves the movement of a DP (*the princess*) into a subject position and therefore has much in common with all the other movements we have discussed. While this movement seems to involve only DPs, the restriction turns out to be tighter than this: only argument DPs can be involved. Thus an adjunct DP is never moved in any of these constructions:

(17) a. The lawyer amended the contract last night.
* Last night was amended the contract.
b. We mashed the potatoes well every time.
* Every time mash the potatoes well.
c. There arrived a package last week.
* Last week arrived a package.
d. It seems that each day John loses his passport.
* Each day seems John to lose his passport.

Moreover, these arguments always move to the subject position, a typical place for arguments. The positions in which arguments tend to appear can be referred to as Argument positions or **A-positions**. 'An A-position is one in which an argument such as a name . . . may appear in D-structure; it is a potential θ-position' (Chomsky, 1981a, p. 47). All the movements seen so far move arguments from one A-position into another and so this kind of movement has been called **A-movement**.

EXERCISE 4.1

Typically the pleonastic subject that appears with a raising verb like *seems* is *it*, while that which appears with an unaccusative is *there*:

It seems that Celia is sewing socks. (* there seems that . . .)
There arrived an architect from Angola. (* it arrived . . .)

However, in the following sentence there is a *there* subject of the raising verb:

There seems to be a problem with the pancakes.

How is this possible?

4.1.2 Ā-movements

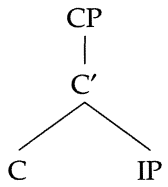
Let us contrast the above movements with another type: wh-movement. The previous chapter described how a wh-element moves to the front of the clause in wh-questions:

(18) a. – the lawyer has amended which contract?

b. Which contract has the lawyer amended –?

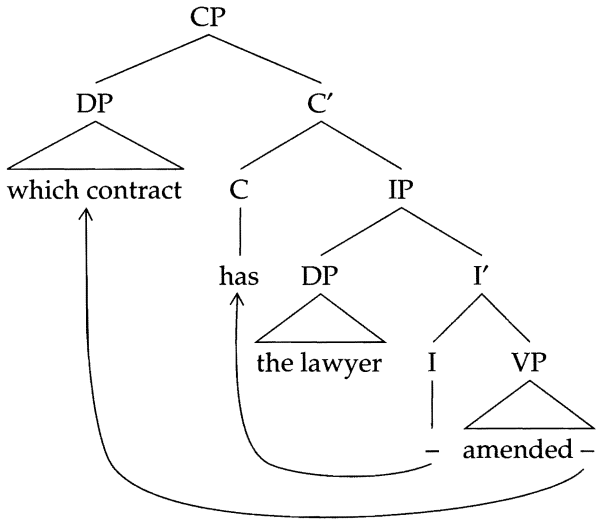
No details were mentioned of the structural position that the wh-element moves to. It is clearly to the left of the subject in the IP specifier position and so presumably outside the IP, i.e. within the domain of the CP above the IP. The earlier discussion of the CP claimed that the CP is headed by a complementizer which takes the IP as its complement:

(19)



Given that the C is a head position and the wh-element is a phrase, we would not expect the wh-element to move to C. However, as yet we have found no use for the specifier of CP. As the specifier position is one where phrases can go, this might well be where the wh-element moves. This solution has the additional advantage of leaving the C position available to accommodate the auxiliary, also fronted by another movement to be discussed later. Thus, a sentence with a fronted wh-element will have the following structure:

(20)



Now that these details are in place, this movement can be compared with others. Although the moved wh-element here happens to be a DP, wh-movement does not always involve DPs. For example APs and PPs can undergo the same type of movement:

(21) a. *How bad* did the news seem?
b. *By whom* was the murder victim discovered?

Moreover, not only does the moved wh-element not have to be a DP, it doesn't even have to be an argument; wh-adjuncts can undergo the same movement:

(22) a. *When* did the maid change the linen?
b. *For how long* was the miscalculation kept from the public?

Thus, this kind of movement seems different from the previous movements which only involved DP arguments. The position to which the wh-element moves also differs from that in A-movement in that the specifier of the CP is not a position where we tend to find arguments and hence it is not an A-position. Those positions that are not A-positions are called **Ā-positions** (the bar on 'A' standing for 'non'); this kind of movement is therefore referred to as **Ā-movement**. 'Other positions we will call "Ā-positions" in particular, the clause-external position occupied by operators such as *who* (Chomsky, 1986a, p. 80).

There are other movements which move elements into Ā-positions and so fall into the same category as wh-movement. English for example has a handful of constructions that go against its standard SVO order by moving elements in front of the Subject. For example both topicalization and negative-fronting (as in *Never again would I fly British Airways*) move elements to some position to the left of the subject (and so out of the IP). But this time the landing site of the movement is to the right of the complementizer. In:

(23) I thought that *this exam*, everyone would pass.

this exam has been moved in front of the subject, while in:

(24) I vowed that *never again* would I fly British Airways.

never again has been fronted before the Subject (together with compulsory inversion of the auxiliary *would*).

Some have taken this evidence to support the claim that there is further functional structure built on top of the IP than the CP alone, as mentioned in chapter 3. Rather like the idea of an articulated IP, where various functional projections are built on top of the VP headed by functional heads expressing tense and agreement, an 'articulated CP' could have projections headed by functional heads expressing such notions as topic and focus. These would provide specifier positions for various elements to move to, each being an Ā-position. However, given that the Articulated INFL Hypothesis is still highly debatable, not everyone is willing to accept the Articulated CP Hypothesis either.

EXERCISE 4.2 Many relative clauses are formed by what looks to be the same process utilized in wh-interrogatives:

(i) The man [who you met].
(ii) I asked [who you met].

What arguments can you give that *who* in (i) is moved to the specifier of the CP?

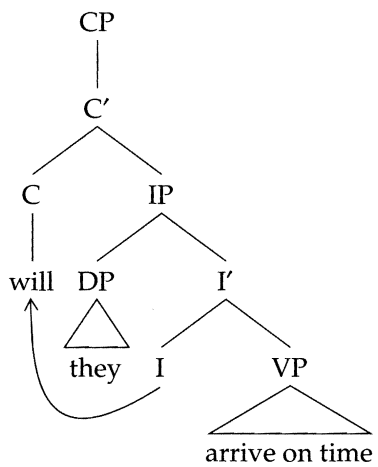
4.1.3 Head movements

The last type of movement to be discussed differs from those seen so far in involving the movement of heads of phrases rather than of the phrases themselves, hence is known as **head movement**. The most obvious kind of head movement is subject–auxiliary inversion, which has already slipped into the discussion without comment. Head movement can be found in certain interrogative clauses, either accompanying wh-movement or by itself in yes-no questions:

(25) a. When will they – arrive?
b. Will they – arrive on time?

As seen in (25), an auxiliary verb such as *will* generated in the I position ends up to the left of the subject *they* in these kinds of interrogatives. Given that the wh-element moves to the specifier of the CP, it is natural to assume that the auxiliary moves to the actual C position:

(26)



In English, movement from I to C is restricted to certain clause types, such as interrogatives, and only involves auxiliary verbs. English main verbs never move to the C position. So:

(27) * Arrived they – on time?

is ungrammatical as it involves movement of the head *arrive*. The grammatical alternative is:

(28) Did they – arrive on time?

where the auxiliary *do* is in C position.

Because the main verb seems stuck inside the VP in English, inversion structures which do not contain an underlying auxiliary verb demonstrate **do-insertion**, involving the insertion of the expletive auxiliary *do* which undergoes the movement to C. In other words, in English, one way of distinguishing a main verb from an auxiliary is whether it can be fronted in interrogatives:

(29) a. May I go?
b. *Go I?
c. Will he arrive?
d. *Arrive he?

leaving some interesting borderline cases in many people's usage:

(30) a. Dare I go?
b. Do I dare to go?
c. Have I to see him?
d. Do I have to see him?

However, languages that allow main verbs to move to C do not require do-insertion. For example, French forms questions in a similar way to English, except that French main verbs behave like auxiliaries by moving to the front:

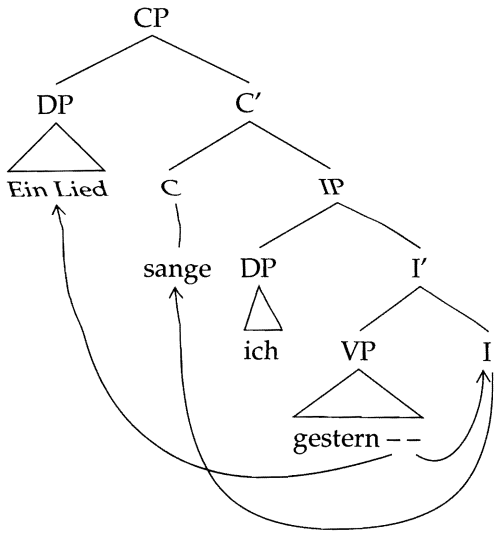
(31) a. Avez-vous chanté dans la classe?
(have you sung in the class)
Have you sung in class?
b. Chantez-vous bien?
(sing you well)
Do you sing well?

In German, it has been argued, *all* finite verbs move to the C position, not just those in interrogatives. This reflects a property common to other languages, such as Scandinavian languages, called *Verb Second (V2)*; some element, whether subject, object or adverbial, is placed in the first position of the clause, followed by the finite verb in the second position:

(32) a. Ich sange ein Lied gestern.
1st 2nd
I sang a song yesterday.
b. Ein Lied sange ich gestern.
1st 2nd
c. Gestern sange ich ein Lied.
1st 2nd

A well-accepted analysis of the German clause has the element at the front of the clause moving to the specifier of the CP and the verb moving to C:

(33)



Support for this analysis comes from the fact that in embedded contexts where the C position is filled with a complementizer, the Verb does not appear in the second position, but stays at the end of the IP, as the German IP is head final:

(34) Ich sage [$_{CP}$ dass, [$_{IP}$ ich ein Lied gestern sange]].
(I said that I a song yesterday sang)
I said that I sang a song yesterday.

Head movement is also involved when Verbs move out of the VP into the I position. In the previous chapter, movement was used to show that verbal inflections are generated as the head of their own phrase and that they become attached to verbs via a movement. In many cases this movement involves the verb. For example, English aspectual auxiliaries *have* and *be* are generated in their own VP projection (as they are not in complementary distribution with modal auxiliaries and so are not generated in I). Their movement to I is shown by the following observations:

(35) a. I should [$_{VP}$ never [$_{VP}$ have met him]].
b. I had [$_{VP}$ never [$_{VP}$ – met him]].

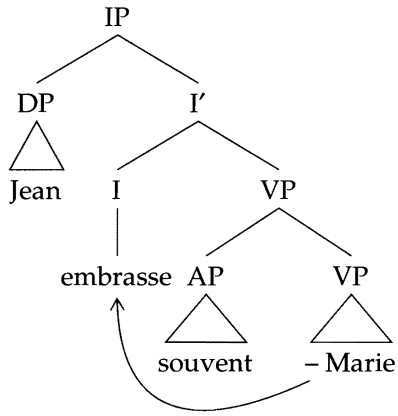
The adverb *never* is presumably adjoined to the left of the VP and, when the I position is filled by a modal auxiliary, the aspectual *have* is inside the VP to the right of the adverb. However, in the absence of a modal, the aspectual auxiliary is finite and sits to the left of the adverb. So the auxiliary moves from within the VP to join with the tense and agreement inflections.

Again, languages differ over which verbs can move from V to I. We have seen that English main verbs cannot move to C, undergoing inversion, though French verbs can. It turns out that English main verbs cannot move to I either, though French verbs can, as we might expect:

(36) Jean embrasse souvent Marie.
(John kisses often Mary)
John often kisses Mary.

Sentence (36) shows that the French main verb *embrasse* precedes the VP adverb *souvent*, stranding its object *Marie* inside the VP by moving to the inflection position.

(37)

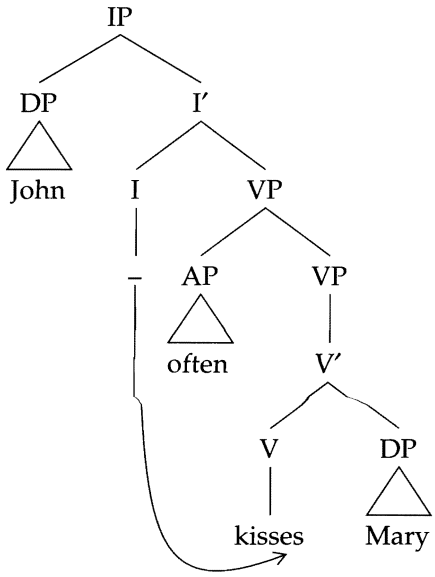


This is not possible in English:

(38) a. * John kisses often Mary.
b. John often kisses Mary.

Instead, the Verb *kisses* remains inside the VP, to the right of the left-adjoined adverb *often*. However it should be noted that the verb is still finite, bearing inflections for tense and agreement. Thus the verb must still get together with the inflections somehow, even though it does not leave the VP. The most obvious solution to this puzzle is to assume that the inflections in this case move down onto the Verb:

(39)



This is more or less the analysis that Chomsky had proposed in 1957, known as *affix hopping* (Chomsky, 1957). This assumed that inflections were generated to the left of the verbal element that bears them at the surface and are then subject to a transformation which shifts them one step to the right:

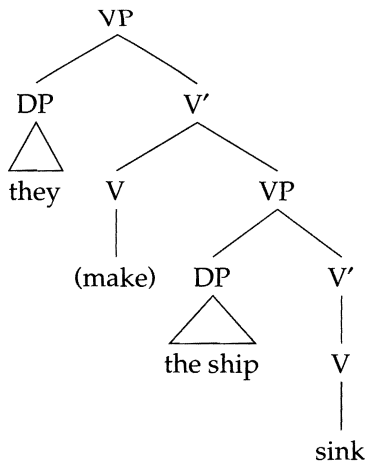
(40) a. John es often kiss Mary.

b. John – often kisses Mary.

The contentiousness of this analysis is that it involves a rightwards and downwards movement, which is not only very unusual (all of the movements we have so far considered are leftwards and upwards), but also seems to violate a number of principles to be discussed later in this chapter.

The final case of head movement to be mentioned involves the movement from one verbal position to another. In the previous chapter our analysis of causative constructions involved an abstract causative verb with a VP complement headed by a thematic verb:

(41)



To get the right word order, it is proposed that, when the causative verb is abstract, the thematic verb moves from its own V position to that of the causative verb. This is rather like V to I movement, especially if the abstract causative verb is taken as a bound morpheme which must attach to the main verb. Indeed, in some languages the causative can be expressed as an overt morpheme which attaches to the verb, as in the following Hungarian example:

(42) János el -gur-it -at -ta a labdát.
(János away-role-cause-past-3.sing. the ball-acc.)
John rolled the ball.

This example shows how the causative morpheme sticks to the end of the verb, as do the tense and agreement morphemes. The morphologically complex verb is created by successive movement of the verb, first to the causative head, then to the tense head and finally to the agreement, picking up the morphemes as it goes:

(43) a. [$_{AgrP}$ ta [$_{TP}$ at [$_{vP}$ it [$_{VP}$ el-gur a labdát]

b. [$_{AgrP}$ ta [$_{TP}$ at [$_{vP}$ el-gur-it [$_{VP}$ – a labdát]

c. [$_{AgrP}$ ta [$_{TP}$ el-gur-it-at [$_{vP}$ – [$_{VP}$ – a labdát]

d. [$_{AgrP}$ el-gur-it-at-ta [$_{TP}$ – [$_{vP}$ – [$_{VP}$ – a labdát]

EXERCISE 4.3

Koopman (1984) notes the following word order patterns in Vata, a Kru language of the Ivory Coast of Africa:

A li saka.
(we eat rice)
We ate rice.

A la saka li.
(we have rice eat)
We have eaten rice.

What is the underlying position of the verb in this language and how can head movement help us to account for the surface word order?

TYPES OF MOVEMENT

A-movement: movement of DPs to argument positions, e.g.
Movement to subject: [$_{IP}$ will [$_{VP}$ the Labour Party win again]]
[$_{IP}$ the Labour Party will [$_{VP}$ – win again]]
Ā-movement: movement of phrases to non-argument positions, e.g.
Movement to specifier of C: [$_{CP}$ [$_{IP}$ they will win the election when]]
[$_{CP}$ when will [$_{IP}$ they win the election –]]
Head movement: movement of heads of phrases rather than whole phrases, e.g. Subject–auxiliary inversion: [$_{CP}$ [$_{IP}$ they will win the election]]
[$_{CP}$ will [$_{IP}$ they – win the election]]

4.2 Further developments to the theory of movement

4.2.1 The Projection Principle and Trace Theory

No details of the restrictions on transformations have so far been provided. Some basic restrictions originated in the 1970s but continued to be assumed in GB Theory. One aimed to curtail the power of transformations so that they could not make

just any kind of change at all to a structure. D-structures should be left intact by a movement transformation, which simply moved things about. This restriction was known as the **Structural Preservation Principle**. Effectively this ensures that the landing sites of a moved element should already be present at D-structure and not be created just for the purposes of the movement itself. This is clearly demonstrated in A-movements, which use the specifier of IP as their landing site. This canonical subject position is obligatorily present in all clauses, as dictated by the Extended Projection Principle (EPP), introduced in the previous chapter. Thus A-movements move NPs into a position which is present, but vacant, at D-structure. We can assume the same is true for other movements we have reviewed. So the specifier of CP, the I and the C position are all potentially present at D-structure, but are vacant so that they can be moved into by the relevant element.

As a consequence of structural preservation, the extraction site of a moved element cannot disappear when it is vacated as this would alter the structure from its D-structure configuration. So extraction sites are left empty at S-structure. This has already been hinted at above by marking the extraction site with a '–'. However, the empty slot created by moving an element out of one position is very different from the empty slot which exists at D-structure and can be filled by something moving into it. Nothing for instance can move into an empty position from which an element has moved, as demonstrated by the following example:

(44) * Will he haven't – read the paper?

In (44) the modal *will* moves to C, vacating the I position. In English, aspectual auxiliaries such as *have* can move from V to I and only in I can the contracted negation *n't* attach to the auxiliary. The ungrammaticality of (44) indicates that the movement of *have* is impossible. The reason is that when an element moves, the position it vacates is not entirely empty but is filled by a **trace** of the moved element. A trace retains the syntactic and semantic properties of the moved element, to which it is linked, but has no phonological realization; it exists as an abstract invisible entity that is never physically present in speech – a kind of invisible place holder. '[W]hen a category is moved by a transformation, it leaves behind an empty category, a "trace"' (Chomsky, 1986a, p. 66). So the trace left behind when a DP moves from object position to subject position in a passive structure will be a DP which shares the same θ-role and referential properties as the moved element, but will not be pronounced:

(45) The paper$_i$ was read t_i.

Here the trace is represented by a *t* which is co-indexed ($_i$) with the moved element to show the link between them. In the previous chapter we encountered the phonologically empty pronouns PRO and *pro*, which share with traces the fact that they are unpronounced. Together these elements are called **empty categories** – elements necessary to the D-structure of the sentence and its interpretation that never occur visibly in the S-structure. 'There is, in fact, substantial evidence of

various sorts to support the hypothesis that empty categories do appear in representations at various syntactic levels' (Chomsky, 1986a, p. 67).

For some purposes traces can be considered independent elements. However, in other ways we might not want to view a trace as a separate entity from the moved element. For one thing, they share the same θ-role as the moved element; yet the Theta Criterion insists that θ-roles can only be assigned to one argument and cannot be shared out. This is solved by the notion of a **chain**. A movement chain consists of the moved element and its trace (in fact, as we will see, an element can move more than once and so there can be more than one trace in a chain). 'A chain is the S-structure reflection of a "history of movement", consisting of the positions through which an element has moved from the A-position it occupied at D-structure' (Chomsky, 1986a, p. 95). So the sentence:

(46) When$_1$ is$_2$ the concert t_2t_1?

has two chains. One:

(47) (*when*$_1$, t_1)

links *when* to its original place at the end of the sentence. The other:

(48) (*is*$_2$, t_2)

links *is* to its original position.

Each chain acts as a single entity so far as principles such as the Theta Criterion (chapter 3, section 3.4) are concerned. Thus this principle can be restated as:

(49) **Theta Criterion**
All theta roles must be assigned to one and only one argument chain.
All argument chains must bear one and only one theta role.

In GB Theory structural preservation was more neatly tied into the way X-bar Theory determined D-structures under the Projection Principle. As seen in the previous chapter, GB Theory states that syntactic levels of representation are projected from the lexicon, in the sense that information in the lexicon determines major aspects of structure. In D-structure the principles of X-bar Theory create category-neutral templates for structures; all the details of particular structures are 'projected' from the lexical items inserted into these templates. Thus the category of the head itself determines the category of the phrase and subcategorization properties of the head determine properties of the complements. That is, the lexical entry:

(50) drive: [V, DP]

for a Verb determines that the phrase to be generated has to be a Verb Phrase and that it has to have an Object; someone drives something:

(51) Sarah drives a Ford.

If S-structures are also linked to lexical information, from which they cannot deviate, it follows that no movement can radically alter a structure: if a verb is transitive at D-structure, it remains transitive even if its object moves at S-structure, as the trace of the object will still remain in the complement position. To take another example, transforming:

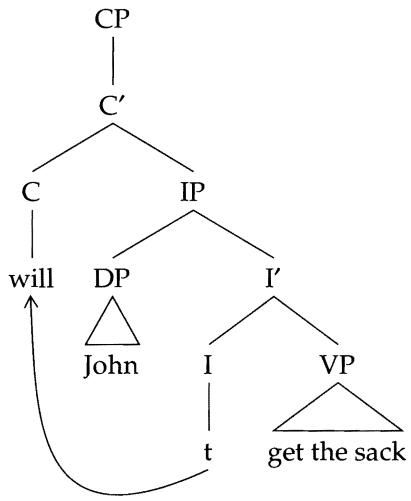
(52) John will get the sack.

to:

(53) Will$_1$ John t$_1$ get the sack?

leaves the trace t in the same position that *will* originally occupied:

(54)



Furthermore, head movements do not change the categorial status of the positions the elements move to. When an inflectional element such as a modal auxiliary moves to C, C does not become I projecting an IP, but remains C projecting a CP. So in (54) the auxiliary *will* is still attached to a C in a CP even if it originated from an I in an IP.

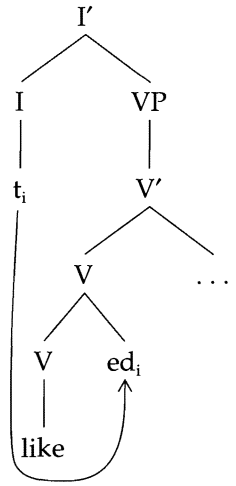
EXERCISE 4.4 In the following sentences insert traces into the appropriate positions and coindex them with the relevant moved element:

They should arrive on time.
John saw Mary.
Who could Henry fight?
This room, what colour would you paint?

4.2.2 *Substitution and adjunction*

The majority of the movements considered so far have involved taking an element from one position to another which is vacant. We can envisage this as filling the empty position with the element undergoing the movement, known as **substitution**: one element simply substitutes for the empty slot by moving into it. However, some movements are not like this. For example, V to I movement, I to V movement and causative V to V movement all move an element to a position which is already occupied, resulting in the formation of a morphologically complex element. This movement is assumed to proceed in the following way: when an English inflection moves to the verb, it 'sticks' itself onto the verb, creating a morphologically complex verb containing both the original verb and the inflection:

(55)



So here the past tense *-ed* has moved from the head of I to the V already filled by *like* and has become joined to it to get the combined element *liked*, in the spoken language getting the [t] pronunciation, in the written language losing an <e>. The structure that is created is essentially an adjunction structure with the inflection adjoined to the Verb. For this reason this kind of movement is known as **adjunction**.

4.2.3 *Move α*

The structure-preserving nature of movement discussed above imposes severe restrictions on what can move where. For example, phrases can only move to phrasal positions and heads can only move to head positions. Ultimately there is no need to state such restrictions as part of the movement rules themselves as they follow from the general Projection Principle. This enables us to simplify the movement rules further, thus carrying out the aim of making more and more general statements rather than structure-specific rules. The restrictions introduced in the following sections allow even more simplification, so that, at the start of the 1980s,

it could be proposed that there was only need of one movement rule **Move α**, where α stands for any element. This simply stated that elements can move about in a tree – 'move any category anywhere' (Chomsky, 1982, p. 15). The details of movements – what actually moves where – were controlled by restrictions placed on movements in general.

The model of GB Theory developed throughout the previous chapter can therefore be modified as follows:

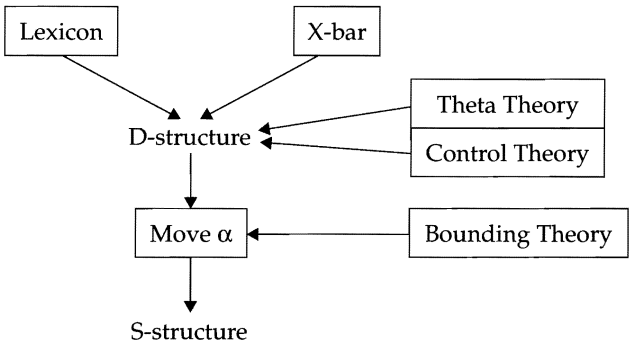


Figure 4.1 Government/Binding Model, amplified from chapter 3

CONCEPTS OF MOVEMENT

The Projection Principle (Chomsky, 1981a, p. 29): Representations at each syntactic level (i.e. D- and S-structure) are projected from the lexicon, in that they observe the subcategorization properties of lexical items.

Trace: the invisible marker left behind in the structure when some element moves

Mary has written a novel.
Has_1 Mary t_1 written a novel?

Chain: an object created by movement consisting of the moved element and all its traces, for example the chain (*where*$_1$, t_1) in the sentence

Where_1 did you put it t_1?

Adjunction: moving an element to another element already in the landing site, with the moved element adjoining to the other:

-ed like → liked

Move α (i.e. move any element anywhere): the maximally generalized transformational rule which simply licenses movement – the details of actual movements themselves are settled by the interaction of other modules of the grammar.

The following sections extend this model by adding yet more modules which interact with Move α in complex ways to provide a detailed analysis of movements.

EXERCISE 4.5

By definition a substitution movement can apply only once, as when an element is moved into a vacant position, it is no longer vacant. Adjunction, on the other hand, can in principle be applied any number of times. Topicalization, such as:

My mother, I always listen to.

has been analysed by some as a substitution movement, moving the topic into the specifier of an abstract 'Top' head of a phrase (TopP) which is part of the articulation of the C-system. Others, however, have claimed it to be an adjunction movement, adjoining the topicalized element to either the IP or the CP. Considering the following data, how many TopPs can there be and where do they come in relation to the CP? Is it better to view topicalization as an adjunction movement?

Today, I will meet my brother in Edinburgh.
Today, in Edinburgh, I will meet my brother.
Today, in Edinburgh, my brother, I will meet.
I said that today, in Edinburgh, my brother, I will meet.
Today, in Edinburgh, this man, who would want to meet?

4.3 Bounding, Barriers and Relativized Minimality

Chapter 2 introduced the notion that grammatical movements are short as a consequence of restrictions imposed directly on movement by the independent module of Bounding Theory. Bounding has a long history in transformational grammar dating back to work in the 1960s by Chomsky (1964) and Ross (1967). We have already mentioned Ross's wh-Island Constraint (p. 71), which formed part of a general programme of research aimed at identifying several constructions which did not allow anything to move out of them, collectively called **islands**. But this approach was very construction specific and hence descriptive rather than explanatory in nature. In the 1970s a more general approach to bounding was introduced, which lasted well into the 1980s, involving a main principle known as **Subjacency**. This section introduces Subjacency and its successors, **Barriers** and **Relativized Minimality**.

4.3.1 Subjacency

The rationale for Subjacency is to prevent movements being overly long by identifying certain **bounding nodes** in a structure that cannot be crossed. A more

general theory can be built in this way instead of by identifying the specific constructions that prevent movement, as the Islands approach attempted to do. However, it is impossible to identify a single node which is never crossed by some movement. Many movements involve shifting an element from one clause to the next, clearly crossing the clausal nodes IP and CP. These would have been obvious nodes to nominate as absolute bounding nodes in order to prevent long-distance movement. Yet, while clause to clause movement is a distinct possibility, it seems to be impossible to move out of two clauses in one go. Recall the restrictions on A- and Ā-movement discussed in the previous chapter (pp. 70–3). Long distances can be achieved by an element moving successively in short hops, leaving traces behind, rather than by moving in one giant leap:

(56) a. $[_{CP}$ Who$_i$ did $[_{IP}$ he think $[_{CP}$ t_i $[_{IP}$ she saw $t_i]]]]$?

b. * $[_{CP}$ Who$_i$ did $[_{IP}$ he ask $[_{CP}$ when $[_{IP}$ she saw $t_i]]]]$?

(57) a. $[_{IP}$ He$_i$ seems $[_{IP}$ t_i to be likely $[_{IP}$ t_i to leave$]]]$.

b. * $[_{IP}$ He$_i$ seems $[_{CP}$ that $[_{IP}$ it is likely $[_{IP}$ t_i to leave$]]]$.

So instead of there being a single node which is an absolute bounding node, preventing anything moving over it, bounding nodes seem to 'gang up' to prevent movement: while one of them can be crossed by a single movement, more than one cannot be crossed in one go. Thus the principle of Subjacency can be stated as follows:

(58) **Subjacency**
No movement can cross more than one bounding node.

Looking at the examples in (56) and (57), it seems that IP can be identified as the bounding node blocking movement across it, as both the ungrammatical movements cross only one CP but two IPs.

Ross identified a number of constructions as islands. For example, a nominal phrase which contains a clause also behaves as an island:

(59) * $[_{CP}$ Who$_i$ did $[_{IP}$ he hear $[_{DP}$ the rumour $[_{CP}$ t_i that $[_{IP}$ she murdered $t_i]]]]]$?

In this example, the wh-element makes two movements: one to the specifier of the noun complement CP, and one to the specifier of the matrix CP. As both of these movements cross only a single IP, the Bounding Theory so far discussed cannot account for the ungrammaticality. To rectify this, we might claim that DP

is also a bounding node, noting that the second movement moves over a DP and an IP, violating Subjacency if both are bounding nodes.

There is some parameterization of what counts as a bounding node in any given language. For example, in Italian, some grammatical sentences appear to violate the wh-Island Constraint:

(60) Tuo fratello, a cui mi domando che storie abbiano raccontato.
(your brother to whom myself ask-1.sing. which stories have-3.pl. told)
* Your brother, to whom I wonder which stories they have told.

In this example, the wh-phrase *a cui* (to whom) has moved out of the 'have told' clause to form a relative clause, despite the fact that this clause begins with another wh-phrase, *che storie* (which stories). As we can see from the translation, this sentence, though grammatical in Italian, is ungrammatical in English. The movement, like other wh-Island violations, crosses two IPs which, as we have established, are bounding nodes for English. The grammaticality of (60) means that IP cannot be a bounding node in Italian, but this does not mean that Italian has no bounding nodes and hence allows unrestricted long-distance movement. For example, Italian, like English, does not allow a movement out of a complex nominal phrase:

(61) * Tuo fratello, a cui temo la possibilità che abbiano raccontato tutto.
(your brother, to whom fear-1.sing. the possibility that have-3.pl. told everything)
* Your brother, to whom I fear the possibility that they have told everything.

Rizzi (1982) has claimed that these data are accounted for by assuming that the bounding nodes for Italian are CP and DP rather than IP and DP: in other words what counts as a bounding node is a parameter of variation between languages. The Italian setting of the Subjacency parameter allows elements to be extracted out of two IPs but not out of two CPs. Thus, a moved wh-element can skip one specifier of CP, allowing a violation of the wh-Island Constraint:

(62) $[_{CP}\ wh_i\ [_{IP} \ldots [_{CP}\ wh\ [_{IP} \ldots t_i]]]]$

The movement in (62) crosses two IPs but only one CP; so it is ungrammatical if IP is a bounding node, as in English, but grammatical if CP is the bounding node, as in Italian.

EXERCISE 4.6

1 Here are some sentences involving movement out of islands. Identify where the bounding nodes are and how many are moved over by each movement and conclude whether Subjacency accounts for all island phenomena:

* Who$_1$ did you meet a man who introduced t_1 to Bill?
* Who$_1$ is the fact t that she met t_1 worrying?
* Who$_1$ is t_1 that she met t worrying?
* Who$_1$ did you see Bill and t_1?
* Who$_1$ is a picture of t_1 on your desk?

2 In Hungarian, wh-elements move to the front of the clause, but cannot normally move out of a clause. So cases of long-distance wh-movement are produced by the use of a dummy wh-element in the scope position and the wh-element at the front of its own clause:

Mit mondott János, kit szeret Mari?
(what said John, who (acc.) likes Mary)
Who did John say Mary likes?

What setting of the Subjacency parameter might produce this strategy of question formation?

SUBJACENCY

The Principle of Subjacency: any movement can cross at most a single bounding node.
Bounding Nodes: these can be CP, IP or DP, subject to parameterization specific to the language.
The Subjacency parameter: this selects the bounding nodes for a particular language, for example English selects IP and DP while Italian selects CP and DP.

4.3.2 *Barriers*

A number of the islands identified by Ross are, however, not accounted for so readily by Subjacency. For example, Ross noted that a clause that sits in subject position is an island:

(63) * [$_{CP}$ Who$_i$ was$_j$ [$_{IP}$ [$_{CP}$ t_i that [$_{IP}$ you met t_i]] t_j unexpected]]?

Here the object within the sentential subject moves first to the CP specifier position of its own clause and then to the specifier of the matrix CP. Both of these movements cross only one bounding node and hence Subjacency does not account for the ungrammaticality. Yet the phenomenon seems part of a wider set of observations which indicate that it is easier to move out of a complement than a subject. For example, a wh-element can be moved out of an object DP, but not out of a subject:

(64) a. $[_{CP}$ Who$_i$ did $[_{IP}$ you draw $[_{DP}$ a picture of $t_i]]]$?
b. * $[_{CP}$ Who$_i$ did $[_{IP}$ $[_{DP}$ a picture of $t_i]$ fall off the wall]]?

In this case the grammatical (64a) causes problems for Subjacency as the movement appears to cross a DP and an IP. Yet these observations fall in line with those which demonstrate it is easier to move out of a complement clause than a clausal subject.

In fact, this can be extended to show that it is easier to move out of a complement construction than out of any other kind of construction. Thus, for example, a wh-element cannot move out of a clause or a DP that functions as an adjunct:

(65) a. * $[_{CP}$ Who$_i$ did $[_{IP}$ he leave $[_{CP}$ because he met $t_i]]]$?
(cf. He left because he met Mary.)
b. * $[_{CP}$ When$_i$ did $[_{IP}$ he meet Mary $[_{DP}$ the day before $t_i]]]$?
(cf. He met Mary the day before Wednesday.)

Whether a construction causes problems for movement depends on where that construction is situated rather than on absolute properties of the construction itself. In this case, both the Island approach and Subjacency are mistaken as they both assume that it is properties of specific constructions which block movements.

In his *Barriers* monograph, Chomsky (1986b) initiated a new approach to bounding by proposing that certain constructions become barriers to movement not because of what they are but because of where they sit in a structure. 'A potential barrier may be exempt from barrierhood by an appropriate relation to a lexical head' (Chomsky, 1986b, p. 12). Given that it is structures other than complements which create problems, Chomsky proposed that complements have a special property which prevents them from being barriers to movement. This special property he called **L-marking**, related in some direct way to a lexical head. Obviously complements are related to lexical heads in that they are selected by them. Thus a complement is an L-marked construction while a subject and an adjunct are not L-marked. It would be simple to claim that constructions which are not L-marked constitute barriers to movement, but unfortunately one non-L-marked element is never a barrier to movement, namely IP. To get round this, Chomsky proposed that something is a **Blocking Category**, i.e. something which is a potential barrier, if it is not L-marked; barriers are then defined as Blocking Categories which are *not* IPs. This is incorporated in the following set of definitions:

(66) **Bounding Principle**
A movement which crosses a barrier is ungrammatical.

(67) A construction is a **barrier** if it is a Blocking Category but not IP.

(68) A construction is a **Blocking Category** if it is not L-marked.

(69) A construction is **L-marked** if it is selected by a lexical head.

This theory, however, is still too simple as it predicts that a wh-element can be extracted directly out of a complement clause, instead of having to move first to the CP specifier position. If this were true then we would not get wh-Island effects.

What is needed is to define any CP as a barrier, unless an element moves into its specifier position. At this point the CP will remain a barrier if it is not L-marked, but will cease to be a barrier if it is. Barrierhood is not an absolute property, but is also defined taking into consideration where the moved element is with respect to it. To achieve this means defining barriers for elements in particular positions and these will not necessarily be barriers for elements in other positions. Chomsky achieves this by making the CP a barrier for an element inside the IP by 'inheritance'. 'Let us suppose that CP *inherits* barrierhood from IP, so that CP will be a barrier for something within IP but not for something in the pre-IP position' (Chomsky, 1986b, p. 12). Simply put, the CP inherits its barrier status from the IP which is a Blocking Category for an element that it contains. So a CP can become a barrier, even if it is L-marked, by dominating a Blocking Category for a moving element. An element inside IP would then be allowed to move out of the IP, as this is a Blocking Category but not a barrier, but it could not move out of the CP. Therefore only movement to the specifier of the CP is possible. Once the element has moved out of IP, the IP is no longer relevant for it, and hence the CP ceases to inherit its barrier status. From this point on the element can move out of the CP as long as this is L-marked. We can schematize this in the following way, where italics indicate Blocking Category status, bold face indicates barrier status and arrows represent L-marking

(70) a. He thinks $[_{CP}$ $[_{IP}$ she likes who]]?

b. He thinks $[_{CP}$ who$_i$ $[_{IP}$ she likes $t_i]]$?

c. *Who$_i$ does he think $[_{CP}$ t_i $[_{IP}$ she likes $t_i]]$?*

BARRIERS

Bounding Principle: A movement which crosses a barrier is ungrammatical.
Barrier: A construction is a barrier for an element α if it is a Blocking Category for α (not IP) or it dominates a Blocking Category for α.
Blocking Category: A construction is a Blocking Category for α if it contains α and it is not L-marked.
L-marking: A construction is L-marked if it is subcategorized by a lexical head.
Gloss: A barrier is a node on a tree that serves to block movement. A node is defined as a barrier in relationship to a particular element that it contains and by virtue of its position in a sentence. First a set of potential barriers (Blocking Categories) are defined as those categories containing the element to be moved which are not selected by a lexical category (L-marked). The actual barriers are the non-IP Blocking Categories, or anything that contains a Blocking Category. When an element moves, categories which may have been barriers for it in its original position may not be barriers for it in its new position. Hence long-distance movement as a series of short hops will be possible.

EXERCISE 4.7

Using the definitions given above say whether the bold elements in the structures below count as a barrier for an element in the position marked X:

believe [$_{CP}$ that [$_{IP}$ . . . X . . .]]
believe [$_{CP}$ that [$_{IP}$ . . . X . . .]]
believe [$_{CP}$ X that [$_{IP}$ . . .]]
left [$_{CP}$ because [$_{IP}$ . . . X . . .]]
left [$_{CP}$ X because [$_{IP}$ . . .]]

4.3.3 *Relativized Minimality*

Clearly the Barriers framework is very complicated. Despite its importance, it was not long before simpler approaches to boundedness were proposed. The most influential was Rizzi's Relativized Minimality approach (Rizzi 1990). Like Barriers, this theory of boundedness does not set up absolute barricades for movement, but relativizes the conditions in which movement is blocked to whatever it is that is moving. The general idea is simple: all movements should be to the nearest relevant position, where what constitutes a relevant position depends on the moved element. If it is a head that is moving, then the nearest relevant position is the nearest head position. If an argument is undergoing A-movement, then the nearest relevant position is the nearest A-position; and if an element is undergoing Ā-movement, then the nearest relevant position is the nearest Ā-position, as seen in the following examples:

(71) a. [$_{CP}$ Could$_i$ [$_{IP}$ he t$_i$ [$_{VP}$ have seen the message]]]?
b. * [$_{CP}$ Have$_i$ [$_{IP}$ he could [$_{VP}$ t$_i$ seen the message]]]?
c. [$_{CP}$ Had$_i$ [$_{IP}$ he t$_i$ [$_{VP}$ t$_i$ seen the message]]]?

(72) a. [$_{CP}$ What$_i$ did [$_{IP}$ you think [$_{CP}$ t$_i$ [$_{IP}$ he saw t$_i$]]]]?
b. * [$_{CP}$ What$_i$ did [$_{IP}$ you ask [$_{CP}$ where [$_{IP}$ he saw t$_i$]]]]?

(73) a. [$_{IP}$ He$_i$ seemed [$_{IP}$ t$_i$ to be likely [$_{IP}$ t$_i$ to win]]]?
b. * [$_{IP}$ He$_i$ seemed [$_{IP}$ it is likely [$_{IP}$ t$_i$ to win]]]?

Sentence (71) shows the effects of bounding on head movement. The inflectional modal auxiliary can move to C as C is the next head position up from I. The aspectual auxiliary *have* cannot move to C, however, as this entails moving over the I which is nearer to the auxiliary but filled by the modal. Only if there is no modal, allowing the aspectual auxiliary to move first to I, can this head then move to C.

In the early 1980s GB Theory these observations were accounted for by the **Head Movement Constraint** (Travis, 1984), which stated that heads cannot move over the top of other heads. If we compare this situation to those in sentences (72) and (73), which represent familiar bounding effects for A- and Ā-movements, we can see a similar effect: a wh-element can move to a higher specifier of CP as long as it does not move over another wh-element in a lower CP specifier, and a subject

can undergo raising as long as it is not raised over another subject position. Rizzi's insight was to see these restrictions as instances of one general restriction:

(74) **Relativized Minimality**
No X can move over another X, where X is either a head, an argument or a wh-element.

This simpler restriction on movement has been very influential for research to the present day. Yet it is not entirely obvious how it can account for the observations which motivated the Barriers approach, namely that complements are easier to move out of than non-complements. Indeed, within GB Theory there was no one approach to bounding that could straightforwardly account for all boundedness restrictions on movement. We will see, however, that many of the ideas discussed in this section *re-emerge in more current treatments.*

RELATIVIZED MINIMALITY

An element must move to the nearest relevant position, defined in relation to what movement is involved: head movement, A-movement or Ā-movement.

4.4 Case Theory

The previous chapter introduced Case phenomena but gave no details of the principles which govern them. Case Theory is in fact a central part of GB Theory which plays an important role in the analysis of certain movements, as we shall see after some other principles that belong to this module have been introduced.

4.4.1 Abstract and morphological Case

It is important to establish first what phenomena are covered by Case. Traditionally the term Case refers to the form that nominals take, say in Latin *mensa, mensam, mensas* etc. (table), often depending on their function in a sentence. For example, in Hungarian the subject noun is typically in the nominative form which is unmarked, but the object noun is in the Accusative form marked by a *-t* affix:

(75) a. János elment.
(John-nom. away-went)
John left.
b. Látom Jánost.
(see-1.sing. John-acc.)
I see John.

As we see from the translations, English does not mark Case distinctions on its nominals: where Hungarian has two forms, *János* and *Jánost*, the same form of the noun *John* appears in both subject and object position in English. Only in pronouns is there a formal difference between subjects and objects in English:

(76) He/she/they admired him/her/them.

Thus English rarely marks Case distinctions morphologically.

Nevertheless this does not mean that the notion of Case is redundant for languages without morphological Case distinctions, as Case plays an important role in many languages in determining the distribution of DPs, not just the form of nominals. 'In some languages, Case is morphologically realized, in others not, but we assume that it is assigned in a uniform way whether morphologically realized or not' (Chomsky, 1986a, p. 76). We therefore need a more general notion of Case distinct from the traditional notion of nominal morphological form, to be called **abstract Case** (as opposed to morphological Case). Abstract Case is a property which is borne by a nominal element as a result of occupying certain positions. Languages vary as to whether abstract Case actually has morphological consequences; nevertheless its presence is universal. Case Theory is the module in the GB grammar which addresses issues of abstract Case.

4.4.2 *The principles of Case Theory*

One of the main principles of Case Theory is the **Case Filter**, briefly mentioned in the previous chapter. This ensures that all overt DPs occupy positions to which Case is assigned – bear in mind that PRO does not sit in a Case position:

(77) **Case Filter**
'Every phonetically realized [DP] must be assigned (abstract) Case' (Chomsky, 1986a, p. 74).

This naturally raises the question of what determines which is a Case position and which is not. The answer lies in the principles of Case assignment. The first thing to note is that certain elements are assumed to have the ability to assign certain Cases. The most straightforward instance of this is Accusative Case:

(78) John invited them.

As Accusative Case is associated with objects, the most obvious assigner of this Case is the verb. This is supported by the fact that in some languages not all verbs take Accusative objects. For example, in German the objects of some verbs have dative or Genitive Case:

(79) a. Sie hilft ihm.
(she helped him-dat.)
She helped him.

b. Er konnte sich des Lachens nicht enthalten.
(he could himself the-gen. laughter not refrain)
He couldn't stop himself from laughing.

As the Case of the object depends on which verb is used, the verb is the ultimate source of the object's Case. In English, all verbs assign Accusative Case and in fact Accusative is the usual Case for verbs to assign in most languages. Accusative Case is known as a **structural Case**, as it is generally assigned to the structural position of the verbal object. Dative and genitive objects depend on particular verbs and are said to bear **inherent Case**, which differs in a number of ways from structural Case:

> We distinguish the 'structural Cases', objective and nominative, assigned in terms of S-structure position, from the 'inherent Cases' assigned at D-structure. . . . Inherent Case is associated with θ-marking while structural Case is not, as we should expect for processes that apply at D-structure and S-structure, respectively. Thus, we assume that inherent Case is assigned by α to [DP] if and only if α θ-marks [DP], while structural Case is assigned independently of θ-marking. (Chomsky, 1986a, p. 193)

The objects of prepositions can also be Accusative:

(80) I sent a letter to *him*.

though in some languages some prepositions may have dative or genitive objects:

(81) a. Ich habe kein geld bei mir.
(I have no money with me-dat.)
I have no money with me.
b. Das Dorf liegt diesseits des Flußes.
(the village lies on-this-side-of the-gen. river)
The village lies on this side of the river.

Prepositions can then be treated in a similar way to verbs, as assigners of structural Accusative Case, or, in some languages, as inherent Case assigners.

Nominative Case is typically assigned to subjects. However, not all subjects have Nominative Case. For one thing, PRO sits in the subject position of non-finite clauses and this DP is considered not to bear Case. When infinitival subjects are overt, they are in the Accusative:

(82) a. For *him* to leave now would be unacceptable.
b. I believe *him* to be rich.

All examples of Accusative subjects involve non-finite clauses; it is then the finite clause which has nominative subjects. So the Nominative Case assigner must have something to do with the finite inflections for tense and agreement. It is often claimed that agreement is responsible for Nominative Case assignment on the basis

that subjects of Portuguese infinitives are nominative when the infinitive bears agreement morphology as noted by Raposo (1987):

(83) Será difícil eles aprovarem a proposta.
(will-be difficult they to-approve-3.pl. the proposal)
It will be difficult for them to approve the proposal.

From these observations we can conclude the following about Case assigners:

(84) a. Verbs and prepositions assign Accusative Case.
b. Agreement assigns Nominative Case.

The next question to arise is where these Case assigners assign their Cases to. There are strict restrictions on Case assignment. Just because a verb, for example, has an Accusative Case to assign does not mean that it can assign it to any DP in a structure. Typically verbs and prepositions assign their Cases to their DP complements, and so the restrictions on Case assignment are similar to those on theta-role assignment discussed in the previous chapter (section 3.4). However, the restrictions on Case assignment are slightly looser than those on θ-roles as Case may be assigned to DPs that are not arguments of the Case assigner, under special circumstances. For one thing, Case is not always assigned by a thematic element. Agreement is a functional element, yet it is responsible for assigning Nominative Case to the subject. Other functional elements also assign Case. Consider the Accusative subject in (82a), repeated as:

(85) For *him* to leave now would be unacceptable.

As there is no agreement element on the non-finite inflection, it is assumed not to be a Case assigner. This clause begins with a *for* complementizer, in the absence of which the sentence would be ungrammatical:

(86) * Him to leave now would be unacceptable.

If sentence (86) is ungrammatical because the infinitive subject lacks Case and hence violates the Case Filter, we must attribute the grammaticality of (85) to the ability of the complementizer *for* to Case-mark the subject. In addition this complementizer *for* has the form of a preposition (indeed is often called the *prepositional complementizer*) and that it assigns Accusative Case is therefore not surprising. Thus this complementizer is another functional category with the ability to assign Case.

A further difference between Case and θ-role assignment is that Case can be assigned by a Verb to a DP which is the argument of another predicate. Consider the Accusative subject in (82b), repeated as:

(87) I believe *him* to be rich.

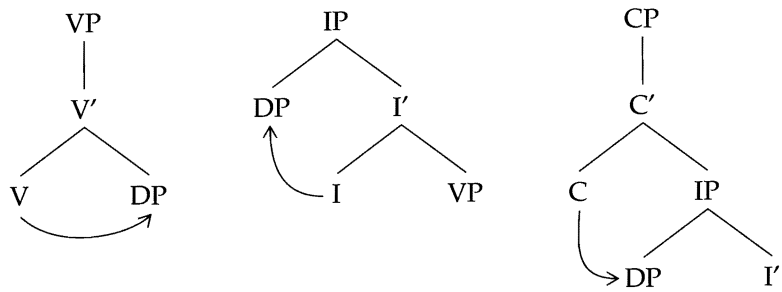
Clearly this *him* is thematically related to the *be rich* predicate and as such can be taken as the subject of the infinitival clause *him to be rich*. As we have said, the

non-finite inflection cannot assign Case and so the Accusative Case borne by the subject must come from elsewhere; the obvious choice is the Verb *believe*, as there is no *for* complementizer. If true, this is radically different from θ-role assignment as no predicate can assign a θ-role to another predicate's argument.

Examining the instances of Case assignment discussed so far reveals at least three different configurations within which Case is assigned:

- from a Verb or Preposition to its complement
- from a finite inflection to its specifier and
- from a complementizer to the specifier of its IP complement (setting aside *believe* for the moment):

(88)



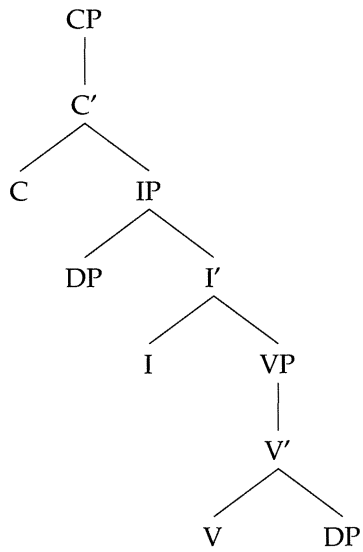
The structural notion of **government** (chapter 3, section 3.6) was construed so as to capture these three cases of Case assignment. If Case is assigned under government, it follows that in the above configurations the Case assigners govern the positions to which Case is assigned. One way in which government resembles θ-role assignment is that it appears to be a relationship between a head and a related phrase. Thus, the set of governors is selected from the set of heads. Not all heads are governors, however. Recall that PRO sits only in ungoverned positions, yet we find this empty pronoun in the specifier of non-finite IPs. It follows that the non-finite I is not a governor. The definition of possible governors is then:

(89) α is a governor if $\alpha = X^0$ (not non-finite I).

Although government is not as restrictive as sisterhood, which is the condition relevant for θ-role assignment, it is still very local. For one thing the governors govern elements that are within their own maximal projections and cannot govern beyond this. To account for this, it is useful to define a structural relation which holds between elements within the same maximal projection. This relationship is often called **m-command**, related to the c-command seen in chapter 3 (p. 114):

(90) **m-command**
α m-commands β if the first maximal projection dominating α also dominates β.

(91)



In the tree in (91), the I m-commands everything inside the IP but not anything immediately in the CP, i.e. the complementizer or its specifier. The subject DP also m-commands everything inside the IP. The Verb m-commands everything inside the VP (the object), but nothing higher.

In terms of locality it is also important that the governed element is not too deeply embedded within the maximal projection of the governor. One way to achieve this is to propose that government cannot cross certain domains. For example, if maximal projections block government, then a governor will be able to govern its own complement and specifier, but not to govern any element contained within these. This would be too strong, however, as we want the complementizer to be able to govern the specifier of its complement. Note that this complement is an IP and we have already seen that IP is exceptional in not counting as a barrier. This suggests that the notion of 'barrier' could be unified for the purposes of bounding and government. If there was just one definition of a barrier which blocked both movement and government relationships, then our theory would be simplified. Indeed, this was one of the aims of Chomsky (1986b) in proposing the notion of a barrier in the first place. From this perspective we can define government as follows:

(92) **Government (Barriers version)**
α governs β if:
(i) α is a governor
(ii) α m-commands β
(iii) there is no barrier between α and β.

Rizzi's notion of Relativized Minimality also attempted to unify locality restrictions on movement and government, but this followed a different tradition of viewing government, dating back to Stowell (1981), in which government is a unique relationship holding between two close elements so that if X governs Y

then Z cannot govern Y if Z is further from Y than X is. Diagrammatically, we can represent this situation thus:

(93) Z X Y

So it is the presence of X, the nearer governor, that blocks the possibility of Z governing Y. This fits into the Relativized Minimality pattern as, if X and Z are potential governors for Y, they will both be of the same type. Viewed from the point of view of movement, an element cannot move from Y to Z over the top of X, where X and Z are the same type, and from the point of view of government, something cannot govern from Z to Y over the top of X, where X and Z are of the same type. In movement, Z and X are potential landing sites, and Y must move to the nearer. For government, Z and X are potential governors and Y can only be governed by the nearer. This gives us the following definition of government:

(94) **Government (Relativized Minimality version)**
α governs β if:
(i) α is a governor
(ii) α m-commands β
(iii) there is no closer governor to β than α.

EXERCISE 4.8

1 Which Cases are assigned to the bold DPs in the following sentences? (Hint: what form of the pronoun replaces them?) Are the results all totally expected and were there any difficult cases? Can you explain why it is difficult to determine which Case is assigned in these cases?

I know that **John** is hiding.
John, I really don't understand.
I expect **John** has left.
I expect **John** to leave.
I never wanted **John** upset.
John smoking is unpleasant.
Q: Who left? A: **John**.
There's **a fly** in my soup.

2 The text has briefly alluded to Genitive Case, which is borne by DP possessors:

John's (his) bodyguard

What could assign this case? Is the observation that some DPs can have a PRO possessor problematic for the treatment of the assignment of Genitive Case under the assumptions made above?:

John's smoking is unpleasant. PRO smoking is unpleasant.

CASE THEORY AND GOVERNMENT

'Case Theory deals with assignment of abstract Case and its morphological realisation' (Chomsky, 1981a, p. 6).
Case Filter: all overt DPs must be assigned Case.
Case is assigned to all DPs by case assigners:

- Nominative is assigned by agreement:

  He disappeared.

- Accusative is assigned by the Verb or Preposition:

  I liked him.
  She gave the book to him.

Case is assigned under government, which means that the Case assigner m-commands the DP that it assigns Case to and:

(i) there is no barrier between the two (Barriers version) or
(ii) there is no closer governor of the DP (Relativized Minimality version).

4.4.3 *Exceptional Case-Marking*

Let us briefly return to the case of the Verb *believe*, which takes an infinitival clause complement with an Accusative subject.

(95) I believe him to be rich.

As this infinitival clause is not introduced by a *for* complementizer, the Accusative Case cannot come from the C-position. In fact, this clause cannot be introduced by a complementizer:

(96) * I believe for him to be rich.

So these kinds of clauses have no C-system and verbs like *believe* can take IP complements. Such verbs are known as **exceptional verbs** precisely because they are exceptions in being able to have IP non-finite complements. Given that IP is not a barrier to government, it follows that exceptional verbs can Case-mark the subjects of their IP complements, in exactly the same way that the *for* complementizer Case marks this subject. Such constructions are called **Exceptional Case-Marking** constructions, or ECM constructions for short.

> In general, a verb selects a full clause [CP] not [IP]; [CP] not [IP] is the normal canonical structural realisation of proposition. Thus *try*, not *believe*, illustrates the general case; such examples [involving *believe*] are often called 'exceptional Case-marking' constructions. In languages very much like English (French and German, for example) these constructions do not exist, and the counterpart of *believe* behaves like *try* in English in this respect. (Chomsky, 1986a, p. 190)

4.4.4 Of-*insertion*

As mentioned in chapter 3, according to one view, nouns and adjectives are not Case assigners and thus cannot have bare DP objects:

(97) a. * a picture George
b. * very fond his mother

Overcoming this problem involves inserting an expletive preposition, which has the role of assigning Accusative Case to these objects:

(98) a. a picture *of* George
b. very fond *of* his mother

Chomsky (1986a) was concerned about the repercussions of this theory, however, and wondered why an *of* could not be inserted to overcome the Case Filter in other cases in which DPs occupied Caseless positions, say the subjects of non-finite clauses:

(99) * of John to leave now would be rude

Moreover, we do not get *of*-insertion on the subject of a non-finite complement of a noun formed from an exceptional verb:

(100) a. I believe him to be rich.
b. * my belief him to be rich
c. * my belief of him to be rich

If Nouns do not assign Case, then we can account for the ungrammaticality of (101b), as the nominalized verb would not be able to Case-mark the subject. But again, if *of*-insertion is a way to circumnavigate the Case Filter in these cases, why should it not be possible in (100c)? Chomsky's answer to these problems was to suggest that Nouns and Adjectives *do* assign Case.

> Suppose we revise . . . Case theory . . . , regarding nouns and adjectives as Case-assigners along with verbs and prepositions. We distinguish the 'structural Cases' objective and nominative, assigned in terms of S-structure position, from 'inherent Cases' assigned at D-structure. The latter include oblique Case assigned by prepositions and now also Genitive Case, which we assume to be assigned by nouns and adjectives just as verbs normally assign objective Case. (Chomsky, 1986a, p. 193)

Of is simply the realization of the Genitive Case that is assigned by these heads and therefore it is not free to appear anywhere we find a Caseless DP as the theory of *of*-insertion would suggest.

This would still not account for the ungrammaticality of (100c), however, as, if the noun is able to assign Genitive Case to its object, it is unclear why it cannot do so to the subject of the non-finite IP in an ECM construction. Chomsky's solution is that genitive and Accusative Cases differ in that the former is an inherent Case while the latter is a structural Case. As we have mentioned, structural Cases are assigned to elements which occupy certain structural positions: i.e. those governed by Case assigners. Inherent Cases, on the other hand, seem more restricted in that they are only assigned to specific arguments and as such the assignment of inherent Case is much more like the assignment of θ-roles. If this is the case then inherent Cases cannot be assigned to DPs which are not arguments of the Case assigner, and hence we cannot get ECM constructions involving the assignment of inherent Case.

INHERENT AND STRUCTURAL CASE

Structural Case is assigned by a Case assigner (Verb, Agreement) to the position that it governs and thus any DP which occupies this position will receive the assigned Case no matter what its thematic relationship to the Case assigner.

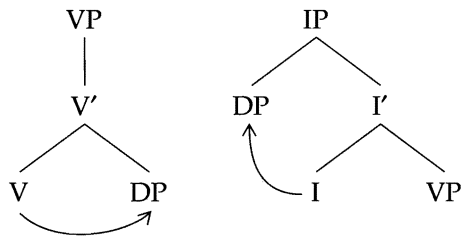
Inherent Case is assigned by a particular Case assigners to a lexically specified argument and therefore it can only be assigned to a DP which is thematically related to the Case assigner.

4.4.5 *Adjacency and conditions of Case assignment*

One last issue remains to be discussed concerning the basic principles governing Case assignment. The original idea, on which the above discussion has been based, was that all Cases are assigned under the same condition, namely government. However, there are differences between Accusative and Nominative Case assignments.

The most obvious concerns the direction that these Cases are assigned in: verbs and prepositions assign Accusative Case to their complements which are to their right, while agreement assigns Nominative Case to its specifier which is to its left:

(101)



It turns out that the direction in which Cases are assigned in is parametric and does not depend on the positions that DP arguments happen to occupy according to the principles of Theta Theory. For example, Koopman (1984) analyses Chinese as having a head-final VP, as most complements and adjuncts tend to precede the verb. However, Chinese objects tend to follow the verb ('aspect' is abbreviated to 'asp.'):

(102) Zhangsan zuotian zai xuexiao kanjian-le Lisi.
(Zhangsan yesterday at school saw-asp. Lisi)
Zhangsan saw Lisi at school yesterday.

Koopman accounts for this by assuming that Chinese objects are generated in front of the verb, but then move behind it, as the verb assigns its Case to the right. If Chinese assigned Case to the left, then the object would not have to move and the sentence would have SOV word order. Given that this word order is possible in many languages, it follows that languages differ with respect to the direction in which Accusative Case is assigned. English obviously chooses to assign Accusative Case, and Genitive Case, if nouns and adjectives assign this, to the right. But, from this point of view, Nominative Case is exceptional in being assigned to the left.

Another difference between nominative and Accusative Case can be seen in the following observations. Generally PP complements can come in any order, but DP complements must come immediately after the verb. Consider the following contrast:

(103) a. I spoke to the students about the assignment.
b. I spoke about the assignment to the students.

(104) a. He put the book on the shelf.
b. * He put on the shelf the book.

This contrast can be accounted for in terms of a restriction on Case assignment. If Case assignment is not only limited structurally, under the notion of government, but also is limited *linearly*, i.e. the Case assigner must be adjacent to the DP that it Case-marks, then it follows that PP complements do not have to be adjacent to their predicate but DP complements do. This is known as the principle of **Adjacency**. Adjacency is also at work in the following examples:

(105) a. * We were anxious for tomorrow him to arrive early.
b. * I believe sincerely him to be rich.

Note that with finite complement clauses the complementizer or the governing verb do not have to be adjacent to the subject:

(106) a. We were anxious that tomorrow he should arrive early.
b. I believe sincerely that he is rich.

The difference lies in the fact that with the finite clause the subject is assigned Case by the inflection, not by the verb or the complementizer. Therefore there is no Adjacency requirement between these elements and the subject. In (105),

however, the subject is dependent on the complementizer and the verb, respectively, for its Case; hence it must be adjacent to these elements.

The Adjacency requirement is parametric since not all languages insist that objects be adjacent to their verbs, as the following Hungarian example shows:

(107) Láttam tegnap Jánost.
(saw-1.sing. yesterday John-acc.)
I saw John yesterday.

If Nominative Case is assigned from the finite inflection, this differs from Accusative Case assignment as the inflection and the nominative subject do not have to be adjacent:

(108) He obviously will correct the error.

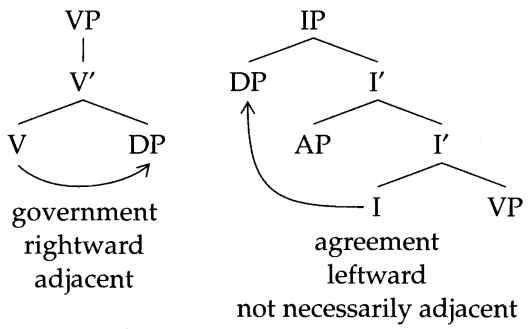
In this case, the finite inflection (*will*) is separated from the subject to which it assigns Nominative Case by the adverb *obviously*.

To account for the differences between nominative and Accusative Case assignment, it has been proposed that they are assigned under different conditions: Accusative Case is assigned under government, as we have described, but Nominative Case is assigned under the condition of **specifier–head (spec–head) agreement** – a relationship assumed to hold between all heads and specifiers which results in both elements sharing features. Morphologically this relationship is overtly realized in cases of subject agreement, where the subject and the agreement head obviously have matching features.

(109) a. A lányok elmentek.
(the girls away-went-3.pl.)
The girls left.
b. Én elmentem.
(I away-went-1.sing.)
I left.

We can also see it more abstractly in wh-movement, where the complementizer head shares the interrogative feature of the wh-element in its specifier position, thus determining the whole CP as an interrogative clause. In English, Case assigned under government is oriented to the right and is subject to adjacency, while Case assigned under spec–head agreement is oriented to the left and is not subject to Adjacency:

(110)



This then accounts for the differences between nominative and Accusative Case assignment.

ADJACENCY

Some languages such as English require case assigners to be adjacent to the DP that receives Case:

I liked him very much. *versus* *I liked very much him.

Others, such as French, have no such requirement:

J'aime beaucoup la France. (I like very much France.)

Let us recapitulate the principles of Case Theory before moving on to the role of Case in the analysis of movement. Perhaps the most basic is the Case Filter which determines that all overt DPs must have Case. This universal principle governs the distribution of DPs in all languages, even though they may differ in terms of the amount of Case that is actually realized morphologically. We then define a set of Case assigners, with verbs and prepositions generally being seen as Accusative Case assigners and finite inflections, or more specifically the agreement element, being the Nominative Case assigner. Languages may differ in terms of whether they have inherent Case-assigning elements and, if so, which of these assign which Case. Since these differences concern the properties of individual verbs or prepositions, they are lexically determined. Finally come the general principles governing how Case is assigned. Some cases are assigned under the restriction of government and may be left- or right-oriented. This will determine the position of the DP with respect to the Case assigner. For example, a Verb Object language assigns Accusative Case from the verb towards its right while an Object Verb language assigns Accusative Case to the left. This kind of Case assignment may or may not be subject to the Adjacency requirement. Other Cases, such as nominative, can be assigned under the specifier–head agreement relationship, which may differ in terms of its direction from the parameter setting for the Cases assigned under government and will not require adjacency.

EXERCISE 4.9

1 Are the following constructions ungrammatical for the same reason?:

* John hit he
* a picture him

2 What problems for the Adjacency condition are presented by the following observations?

He criticized severely every proposal the board made.
Whom did you want to see?
That kind of attitude, I really can't put up with.
I gave Sarah the money.

4.4.6 Case and movement

Now we turn to the role Case Theory plays in structures involving movement. The relevance of Case for movement stems from the fact that Case Theory is a module of S-structure and so the principles of Case Theory are irrelevant for D-structure: DPs are positioned only with respect to the requirements placed on them from Theta Theory and the X-bar modules. Some of the positions that these modules require DP arguments to sit in will be positions to which Case is assigned, being governed by a Case-assigning head. Others, however, will not. This will not matter at D-structure as the principles of Case Theory do not apply there. But at S-structure it *will* matter; unless the Caseless DP is moved to a position where it can get Case, the structure will be ruled out as ungrammatical by the Case Filter. Thus Case plays a motivating role for certain movements.

This last statement needs some qualification, however. Saying that Case motivates movement is to speak in metaphors: it is not as though there is some homunculus sitting in the language faculty who moves things around for particular reasons. Movement in GB Theory, viewed as Move α, is something which may or may not take place with no restrictions on the actual operation itself. Some of these movements will produce grammatical structures, conforming to all the principles of other modules, and some will not. Similarly, if a movement does not take place, this may sometimes result in a grammatical structure, sometimes not, depending on whether other principles are satisfied. In the cases under discussion, if the movement does not take place, the result will be ungrammatical as the result is a DP sitting in a Caseless position. On the other hand, if the relevant movement does take place, i.e. moving the DP to a Case position, the result will be a grammatical structure.

Obviously the only movements that Case Theory has anything to do with involve DPs. There is one kind of movement we have discussed above which exclusively concerns DPs: A-movement. Let us consider how Case is involved in this kind of movement. Recall that in raising structures, the subject of the non-finite clause moves to the subject of the raising verb:

(111) a. John$_i$ seems [t$_i$ to like Mary].
b. It seems [John likes Mary].

As we have seen, the subject position of an infinitival clause is generally a Caseless position, unless it is Case-marked by a *for* complementizer or an exceptional verb. As neither of these is present in (111a), the infinitival subject is Caseless and if it were to remain in this position the result would be ungrammatical:

(112) * It seems John to like Mary.

In (111b), on the other hand, the complement clause is finite, and hence its subject receives Case from the finite inflection. Therefore the structure is grammatical with the subject remaining in the complement clause.

Another straightforward instance of Case-motivated movement concerns the movement of the subject out of the specifier of VP, where it originates according to the VP-Internal Subject Hypothesis, into the canonical subject position, the specifier of IP:

(113) [$_{IP}$ He$_i$ will [$_{VP}$ t$_i$ write a letter]].

From what we have said about Case assignment in English, the specifier of VP is obviously a Caseless position: the verb cannot assign its Accusative Case to this position as Accusative Case is assigned under government to the right; the inflection cannot assign its Nominative Case to this position as this is assigned under spec–head agreement to the specifier of the IP. Therefore, if the subject were to remain inside the VP, it would violate the Case Filter at S-structure and the resulting structure would be ungrammatical:

(114) * [$_{IP}$ It will [$_{VP}$ he write a letter]].

In passive structures, the object moves to the subject position and cannot remain in object position:

(115) a. John$_i$ was identified t$_i$.
b. * It was identified John.

We could account for this phenomenon if the object position of a passive verb were Caseless. In this case, the object would be ungrammatical if it did not move; moving it to the subject position where it could receive Nominative Case would overcome the problem. But verbs are generally seen as Accusative Case assigners. Why would a passive verb be different? There are a number of possible answers to this question. We know that when a verb is passivized its argument structure is altered so that it no longer has an external θ-role to assign to its subject. Clearly this change happens in the lexicon before the verb is inserted into the structure. It is possible that at the same time the verb is altered in terms of its Case-assigning abilities: its Accusative Case is **absorbed** as a result of passivization. Another possibility is that passivization alters the category of the verb to something more like an adjective. This is supported by the fact that passive verbs have a very similar distribution to adjectives:

(116) a. John was interrogated. John was tall.
b. the interrogated man the tall man

If adjectives do not assign Case, then the change of category of the verb would explain why passivized verbs do not assign Case. However, this theory does not work under the assumption that adjectives assign Genitive Case, as passive verbs cannot take objects marked as genitive by the preposition *of*:

(117) * It was interrogated of John.

Jaeggli (1986) proposed that the changes that passive verbs undergo do not take place in the lexicon but are syntactic in nature. He suggested that both the external θ-role and the Case of the verb are assigned in passive structures, but not to their usual places. Instead these elements are assigned to the passive morpheme itself.

Whatever the actual analysis of passivization, passive verbs follow a robust generalization concerning verbs and Case assignment. Burzio (1986) noted that verbs which failed to assign an external θ-role also fail to assign structural Case. We

can see this in the case of unaccusative and middle verbs. Again, both of these involve the object obligatorily moving to the subject position:

(118) a. Three men arrived. * It arrived three men.
b. The bread cut easily. * It cut the bread easily.

Both these constructions fail to have external arguments and, given that the object has to move, they both fail to assign Accusative Case to their objects. Burzio's generalization is quite mysterious as it is not at all clear why the assignment of θ-roles and Case is linked in this manner. However, it does seem to hold true for a number of constructions in a wide variety of languages.

An interesting observation concerns the difference between structural and inherent Cases in passive contexts. We have said that inherent Cases are assigned in a similar way to θ-roles in that they are assigned to specific arguments and not to structural positions. That Nominative Case is structural is obvious by the fact that whatever argument sits in the subject position, it will get Nominative Case. Thus an object that moves to the subject position or a subject of a lower clause that raises to a subject position will all bear nominative. If we passivize a verb which assigns an inherent Case, however, it does not lose this Case:

(119) a. Sie hilft ihm.
(she helps him-dat.)
She helps him.
b. Ihm wird geholfen.
(him-dat. was helped)
He was helped.

For this reason it is usually assumed that inherent Case, unlike structural Case, is assigned at D-structure and hence is not affected by movement.

CASE AND MOVEMENT

Case is an S-structure module which means that DPs do not have to occupy Case positions at D-structure. If a DP generated in a Caseless position at D-structure remains there, however, the Case Filter will apply at S-structure and rule the structure ungrammatical. If the offending DP moves from its Caseless position to one that is Case-marked, the Case Filter will be satisfied. This analysis applies to a wide number of structures, including:

subject movement: [John$_1$ will [t$_1$ meet his bank manager]].
passivization: [The bank manager$_1$ will [be met t$_1$]].
unaccusatives: [The bank manager$_1$ [arrived t$_1$ late]].
middles: [The bank manager$_1$ [scared t$_1$ easily]].
subject raising: [John$_1$ seems [t$_1$ to have intimidated the bank manager]].

All of these movements are A-movements and so we can conclude that Case motivation is a property of this type of movement.

EXERCISE 4.10 A-movements used to be called NP-movements (DP-movements, under current assumptions). Why was this?

The Case Theory module can now be added to the GB Model we have been constructing over two chapters.

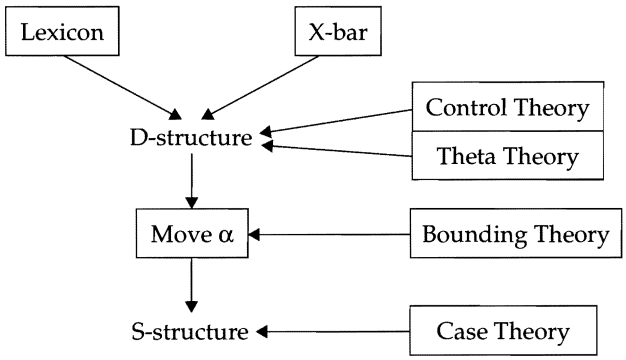


Figure 4.2 Government/Binding Theory (Control Theory added)

4.5 Binding Theory

While the principles of Case Theory determine the S-structure positions of DPs, there are other requirements that certain DPs have to satisfy, which have to do with their referential properties rather than with Case. Take, for example, the following sentences:

(120) a. John said that Bill admires himself.
b. John said that Bill admires him.
c. He said that Bill admires John.

The difference between (120a) and (120b) is merely that the object of the embedded clause is expressed as a reflexive pronoun in the first and a personal pronoun in the second, yet the two sentences cannot mean the same thing. In (120a) the reflexive pronoun must be interpreted as referring to *Bill* and *John* cannot be its antecedent, whereas in (120b) the personal pronoun cannot refer to *Bill* but may take *John* as its antecedent, or alternatively it may refer to someone not mentioned in the sentence. Note, however, that while personal pronouns may have antecedents, in (120c) the pronoun cannot be taken as referring to either *John* or *Bill*. What is demonstrated by these examples is that elements with certain referential properties must appear in structural positions defined with respect to their antecedents. The collected set of principles which govern these phenomena is known as **Binding Theory**. Below we shall introduce these principles and show how they have a regulating role to play with respect to movement.

4.5.1 Binding Theory and overt categories

4.5.1.1 Anaphors, pronominals and r-expressions

Let us start by looking at the referential properties of reflexive pronouns such as *himself, themselves* etc. As seen in (120a), a reflexive pronoun takes a nearby antecedent and cannot refer to something that is too far from it. Thus *Bill* is a possible antecedent but *John* is not. Furthermore, a reflexive pronoun must have an antecedent and cannot refer directly to something not mentioned in the sentence, as personal pronouns can:

(121) a. * Himself is tall.
b. He is tall.

Reflexive pronouns are not the only kind of pronoun which behave like this. Reciprocal pronouns such as *each other* must also have antecedents, which need to be close by:

(122) a. * Each other met.
b. The boys said the girls know each other.

In (122b) the reciprocal pronoun *each other* must refer to *the girls* and cannot be taken as referring to *the boys*. Elements which have these particular referential properties are known as **anaphors**.

Personal pronouns behave differently from anaphors. Not only can they take more distant antecedents, but they do not have to have an antecedent at all:

(123) a. George thinks the girls like him.
b. He left.

The only impossible antecedent for a personal pronoun is one which is too close to it:

(124) George likes him.

In sentence (124), *him* cannot refer to *George* but must be taken as referring to someone not mentioned in the sentence. Personal pronouns therefore have exactly the opposite referential possibilities to anaphors. Elements which behave like this are known as **pronominals**.

Pronouns are not the only elements which have referential properties, however. DPs such as *John, that girl* and *the island just off the coast of France* all refer to something in the world. But such DPs have different referential properties from either anaphors or pronominals in that they can never have an antecedent no matter how far away it is:

(125) He said she likes Hillary.

In this sentence, the DP *Hillary* can be taken as co-referential with neither of the pronominal subjects and must be interpreted as an independent element in the sentence. These elements are called **r-expressions**.

4.5.1.2 The role of indices

When we speak of reference, we are obviously talking about a semantic phenomenon. Yet it is clear that reference plays a role in determining the distribution of certain DPs, and distribution is a syntactic matter. It is well accepted in generative grammar since Chomsky (1957) that syntax and semantics are independent systems, related only by the semantic component interpreting the syntactic representation. It follows therefore that there must be something syntactically represented which is interpreted as reference. In GB Theory, reference is partially represented by the syntactic device of an index. For example, we have seen that a pronominal cannot refer to something which is too close, but it may take an antecedent which is further from it. These two situations can be represented in the following way:

(126) a. John_i thinks Bill_j admires him_i.
b. * John_i thinks Bill_j admires him_j.

What the indices represent is not so much the actual reference of an element as whether two elements are co-referential or disjoint in reference. Thus giving *John* the index *i* and *Bill* the index *j*, we claim that these two elements do not refer to the same entity. More importantly, by giving the pronoun *him* the index *i* or *j* we claim that it is co-referential with any element which bears these indices. Thus, in (126a) the pronoun is co-referential with *John* and in (126b) it is co-referential with *Bill*. Note that these two examples represent two different sentences with different grammatical statuses: the first one is grammatical, the second is not.

The use of indices shows how anaphors and pronominals stand in complementary distribution with each other:

(127) a. John_i thinks Bill_j admires $\text{him}_i/\text{himself}_{*i}$
b. John_i thinks Bill_j admires $\text{him}_{*j}/\text{himself}_j$
c. John_i thinks Bill_j admires $\text{him}_k/\text{himself}_{*k}$

Viewing the examples with different indexing as different structures, in all those structures in which the pronominal is grammatical, the anaphor is ungrammatical in the same structural position and vice versa.

But how do indices come to be part of a structure? One thing is clear: they are not lexically determined: it is not a lexical fact about the pronoun *him* that it must be co-referential with *John*, or any other DP for that matter. Moreover, as indices are structural elements which are interpreted semantically, not semantic elements themselves, they should be created along with the structure. In GB Theory it was proposed that indices are freely assigned to DPs, presumably at D-structure. Each indexation gives rise to a different grammatical entity which is then subject to the relevant principles and deemed grammatical or ungrammatical accordingly.

There is a certain amount of controversy over the use of indices to represent co- and disjoint reference. The main problem is that they are themselves too simple to represent some of the more complex situations. Simple indices can represent the situations in which two DPs refer to exactly the same bit of the world or where they refer to entirely different bits. But it is possible for references to overlap or for one reference to be included in another wider one, as in the following examples:

(128) a. English men generally like cricketers.
b. I think we should leave.

In (128a) the set of *English men* includes some of the set of *cricketers* and vice versa. Thus these two sets have overlapping references which would not be described if they had either the same indices or different ones. In (128b) the reference of *I* is obviously a part of the reference of *we* and again this is captured neither by co-indexation nor by disjoint indexation. The issue is, however, whether or not these considerations play a role in determining grammaticality. If they do not, then there is no point in trying to establish complicated systems of indexing to represent them more accurately. The issues are complex and we will not go into them here, as they tend to be ignored in most literature on binding phenomena. But some have argued that such issues should not be ignored and indeed the assumptions of standard Binding Theory founder exactly on this point.

4.5.1.3 c-command and binding

Let us now turn to the details concerning the referential behaviour of anaphors, pronominals and r-expressions. These are the foundation for the principles of Binding Theory, the relevant module of the grammar.

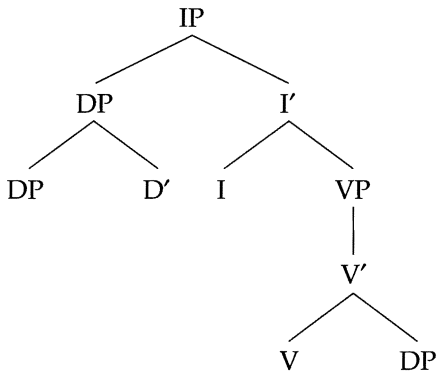
We have already established that anaphors must have close antecedents and that pronominals cannot. However, this needs to be modified in the light of sentences like the following:

(129) a. * John$_i$'s mother likes himself$_i$.
b. John$_i$'s mother likes him$_i$.

The ungrammaticality of (129a) is hard to account for in terms of the distance between the anaphor and its antecedent, as there is little difference in these terms between this structure and the grammatical John$_i$ likes himself$_i$. The main difference lies not in the distance between the anaphor and its antecedent, but in the structural position of the antecedent. In the grammatical case the antecedent is in the subject position and in the ungrammatical case it is in the possessor position within the subject. As a result, the structural relationship holding between the anaphor and the antecedent is different, and in GB Theory it is assumed that this structural relationship is important for the principles of Binding Theory.

A closer look at these structures readily shows the structural relationship which must hold between an anaphor and its antecedent:

(130)



The previous chapter introduced the structural relationship of **c-command**, defined as follows:

(131) **c-command**
An element c-commands its sister and everything dominated by its sister.

From this we can see that the subject DP c-commands the object DP as the subject's sister, the I′, dominates the object. However, the possessor DP inside the subject does not c-command the object which is not included in the D′, the sister to the possessor. Therefore the condition on anaphors is that they must have close *c-commanding* antecedents. The grammaticality of (129b) demonstrates again that pronominals behave in a complementary way to anaphors and cannot have close c-commanding antecedents, though they can have close antecedents that do not c-command them.

It is clear from the above that co-indexation and closeness do not play much of a grammatical role by themselves but that it is the combination of the two that is important. For this reason we define the relationship of **binding** on the basis of these:

(132) **Binding**
α binds β if and only if:
(i) α and β are co-indexed and
(ii) α c-commands β.

4.5.1.4 Binding principles

The structural definition of binding given above allows us to establish easily the principles of Binding Theory that govern the behaviour of referential elements. Let us start with anaphors. These must have a close c-commanding antecedent. An antecedent is by definition an element that the anaphor is co-indexed with and so an anaphor must be bound within a certain local domain, stated as a simple principle of Binding Theory:

(133) **Principle A**
An anaphor must be bound locally.

Now consider pronominals. These cannot have a local c-commanding antecedent and therefore they cannot be bound locally. If we define a notion **free** to mean *not bound*, then the principle governing pronominals can be stated thus:

(134) **Principle B**
A pronominal must be free locally.

The fact that this principle is the exact opposite to Principle A, governing the behaviour of anaphors, accounts for the complementary behaviour of the two.

R-expressions do not have antecedents, locally or not. However, once again, the notion of c-command can be seen as important in determining the allowable relationships between an r-expression and a co-indexed element:

(135) a. His$_i$ own mother distrusts John$_i$.
b. * He$_i$ distrusts John$_i$.

The indexation in (135a) is possible because neither the pronominal nor the r-expression *John* c-commands the other. Therefore there is no binding relationship established here and the structure cannot violate any binding principle. In (135b), on the other hand, the pronominal c-commands and binds the r-expression. R-expressions can then never be bound:

(136) **Principle C**
An r-expression must be free everywhere.

Now we have proposed three very simple principles governing the behaviour of each of the three types of DP. What remains in doubt is what *locally* means in Principles A and B.

4.5.1.5 The governing category

From the examples seen so far, the local domain within which the anaphor must be bound and the pronominal must be free could well be taken as the clause which immediately contains the pronoun. However, this definition turns out to be both too broad and too narrow. Apparently not all clauses count as local domains for binding purposes and, what is more, not all local domains are clauses.

An example of a clause that does not count as a local domain is the following:

(137) John$_i$ believes [himself$_i$ to be discreet].

There are several noteworthy things concerning the embedded clause, which obviously is not a local domain as the anaphor can take an antecedent outside of it. First, the clause is non-finite. Finite clauses, on the other hand, are always local domains:

(138) * John$_i$ believes [himself$_i$ is discreet].

However, it is not the case that non-finite clauses are never local domains, as seen in the following:

(139) * John$_i$ believes [Mary to like himself$_i$].

In this case, *John* cannot be the anaphor's antecedent as it is too far from it, making the non-finite clause a local domain. The difference between (137) and (139) is that in the grammatical case (137) the anaphor is in the subject position. Thus, the non-finite clause does not count as a local domain for its own subject, though it does for its object. Recall that the subject of the non-finite clause is not governed and Case-marked from within the clause, but depends on the exceptional verb for its Case. Thus, unlike the subject of a finite clause and objects in general, the subject of a non-finite clause is governed from outside its clause.

From the above discussion it seems that it is the presence of a governor that is important for marking the local binding domain for a pronoun. For this reason this local domain is often called the **governing category** and may, for now, be defined as follows:

(140) **Governing category**
β is a governing category for α if β is the smallest clause which contains α and the governor of α.

This definition of a governing category allows us to revise the three principles of the Binding Theory:

(141) **Principle A**: An anaphor must be bound within its governing category.
Principle B: A pronominal must be free within its governing category.
Principle C: An r-expression must be free everywhere.

An example of a binding domain which is not a clause, and which will force us to update the definition of the governing category in (140), is as follows:

(142) * Bill$_i$ saw [$_{DP}$ Mary's picture of himself$_i$].

In this example, the anaphor *himself* cannot take *Bill* as its antecedent, despite the fact that both are within the same clause. So there must be a relevant binding domain which is smaller than this clause. The most obvious candidate would be the DP. Note, however, that not all DPs count as binding domains:

(143) Bill$_i$ saw [$_{DP}$ a picture of himself$_i$].

Clearly it is the presence of the possessor that makes the difference and hence this must be included in the definition of the governing category. The possessor of the DP is in many ways like the subject of a sentence: they sit in structurally similar positions – specifier of DP and IP respectively – and the subject is often translated as a possessor in nominalization:

(144) a. *The enemy* destroyed the city.
b. *the enemy's* destruction of the city

For this reason, the possessor is often called the subject of the DP. By the EPP, clauses always have subjects, but DPs only have subjects when there is a possessor. A clause will always be a governing category, but only possessed DPs are. It therefore follows that governing categories are defined by the presence of a subject. The definition of the governing category can be updated:

(145) **Governing category**
β is a governing category for α if β is the smallest constituent with a subject which contains α and the governor of α.

In standard Binding Theory (Chomsky, 1981a), further additions to this definition accounted for more complex data, though we shall not go into these details here. Work on Binding Theory throughout the 1980s tried either to simplify the approach or to rationalize the complexities. For our purposes, however, the above will suffice.

BINDING THEORY

'The theory of binding is concerned with the relations, if any, of anaphors and pronominals to their antecedents' (Chomsky, 1982, p. 6).

Principles of Binding Theory:

Principle A: An anaphor must be bound within its governing category.
Principle B: A pronominal must be free within its governing category.
Principle C: An r-expression is free everywhere.

The governing category:

β is a governing category for α if β is the smallest constituent with a subject which contains α and the governor of α.

So in:

Peter thought himself to be overpaid.

the anaphor *himself* is bound to *Peter* by Principle A; while in:

Peter thought him to be overpaid.

the pronominal *him* must be bound outside the sentence to someone whose name we don't know; and in:

Peter thought the managers were overpaid.

the managers is bound outside the sentence.

EXERCISE 4.11 It has long been known that certain movement phenomena and certain referential phenomena are restricted by similar conditions. Thus, while it is not possible for an anaphor to refer out of a finite clause, it is also not possible for a DP to move out of a finite clause:

* John$_1$ thinks [himself$_1$ is good looking].
* John$_1$ seems [t$_1$ is good looking].

Furthermore, while an anaphor object cannot refer to the subject of a higher clause, movement cannot take place from the object position to the subject of a higher clause:

* John$_1$ believes [Mary to like himself$_1$].
* John$_1$ seems [Bill to like t$_1$].

Under the assumptions of Binding Theory, the facts about the reference of the anaphor follow from Principle A. How can we account for the similarity of the restrictions placed on movement?

4.5.2 *Binding Theory and empty categories*

Having outlined the rudiments of Binding Theory as they apply for overt referential DPs, we now turn to the relationship between Binding Theory and movement. This is made possible by the fact that empty categories, including traces left behind by moved elements, like referential DPs, can be seen as having binding properties. In other words, some empty categories behave like anaphors, some like pronominals, and some like r-expressions; as such these classifications are relevant for all DPs, whether overt or covert. In what follows we will go through each empty category to see how it behaves with respect to the binding principles.

4.5.2.1 *DP traces as anaphors*

It had been noted in the 1970s that the reference of certain pronouns and movement were connected phenomena. Note the similarities in the following sets of data:

(146) a. John$_i$ likes himself$_i$.
b. * John$_i$ believes Mary likes himself$_i$.
c. John$_i$ believes himself$_i$ to be honest.
d. * John$_i$ believes himself$_i$ is honest.

(147) a. John$_i$ was hurt t$_i$.
b * John$_i$ was believed Mary likes t$_i$.
c. John$_i$ was believed t$_i$ to be strong.
d. * John$_i$ was believed t$_i$ likes Mary.

As we see in (146a) and (147a), a reflexive pronoun in object position can take the subject of its own clause as its antecedent and a DP can move from object

position to the subject position of its own clause. This contrasts with (146b) and (147b), which show that it is impossible for an object reflexive to refer to the next highest subject and for a DP object to move to this position. The (c) and (d) examples show further similarities: a reflexive subject can refer to the next subject up from a non-finite clause but not a finite one, and a DP can move from the subject of a non-finite clause to the next subject up but not from a finite clause.

The data in (146) are captured under the Binding Theory presented above. A reflexive pronoun is an anaphor and hence subject to Principle A of the Binding Theory stating that it must be bound in its governing category. For an object, as in (146a) and (146b), the governing category is the minimal clause and as the anaphor is bound within this in the first but not in the second, the first is grammatical and the second ungrammatical. For a subject, as in (146c) and (146d), the governing category differs depending on where the governor is. In a finite clause the governor is the finite inflection of that clause and so this is the governing category. For an exceptional clause, the governor is the exceptional verb, which is outside this clause. Thus the exceptional clause will not be a governing category and its subject can be bound by an element outside it.

The data in (147) can be captured in exactly the same way through one simple assumption: the trace left behind by an A-movement is an anaphor which is bound by the moved DP. This effectively restricts movement to positions within the governing category of the trace and hence this imposes further limitations on this kind of movement over and above those already placed on it by Bounding Theory.

4.5.2.2 *wh-traces as r-expressions*

If all traces were anaphors, the same restrictions would show up for all movements. However, a wh-element can be moved out of a finite clause from either the subject or the object position:

(148) a. Who$_i$ did John think [$_{CP}$ t$_i$ [$_{IP}$ Mary liked t$_i$]]?
b. Who$_i$ did John think [$_{CP}$ t$_i$ [$_{IP}$ t$_i$ liked Mary]]?

Clearly the traces in subject and object positions in the above examples do not behave like anaphors as they are not bound within the finite IP. Indeed, these traces are not bound in the usual way at all. In all the other cases of binding considered so far, the antecedent has been in an A-position. But in (148) the antecedents of these traces sit in the specifier of CP, an Ā-position. Thus, if these traces are bound at all, they are bound in a very different way to anything else we have come across. Furthermore, the following data demonstrate that such traces are not allowed to be bound from an A-position:

(149) a. Who$_i$ [$_{IP}$ t$_i$ thinks [$_{IP}$ he$_i$ should win]]?
b. * Who$_i$ does [$_{IP}$ he$_i$ think [$_{IP}$ t$_i$ should win]]?

In (149a) the trace is bound by the wh-element and the pronominal is bound by the trace. As none of these facts violates any principle of the Binding Theory, the sentence is grammatical and is interpreted as meaning 'who is the person X, such that X thinks that X should win?'. In (149b), however, the trace is bound by the pronominal in the subject position. The meaning of this sentence would be identical to the first and, as this meaning is perfectly well formed, we cannot account

for the unacceptability of this sentence in semantic terms. Given that the pronominal is not bound within its governing category, there is no violation of Principle B and so the ungrammaticality of the sentence must be something to do with the trace. The trace of a wh-element cannot be bound from an A-position, suggesting that these traces behave in many respects like r-expressions.

TRACES AND BINDING THEORY

Traces left by A-movements are **anaphors** and must be bound within the governing category by the moved element. Therefore no A-movement can cross a governing category:

John$_1$ was believed [t$_1$ to like Bartok].
* John$_1$ was believed [t$_1$ liked Bartok].

Traces left behind by Ā-movement are **r-expressions** and therefore cannot be bound by anything in an A-position:

Who$_1$ did he$_2$ say [t$_1$ likes Bartok]?
* Who$_1$ did he$_1$ say [t$_1$ likes Bartok]?

4.5.2.3 Pro *and the PRO Theorem*

So far we have found empty category equivalents to anaphors and r-expressions. This leads us to expect that there must be an empty pronominal, as indeed appears to be true. The empty subject in pro-drop languages, *pro*, behaves very much like an empty personal pronoun. This empty category does not have to have an antecedent, though it can have one as long as this is far enough away, as the following Hungarian sentences show:

(150) a. Pro el-ment.
(away-went-3.sing.)
He left.
b. János$_i$ mond-ta hogy pro$_i$ el-ment.
(John said-3.sing. that away-went-3.sing.)
John said that he left.

This means that there are empty category equivalents to all three types of overt referential elements:

	anaphor	pronominal	r-expression
overt	reflexive	personal pronoun	proper noun
covert	A-trace	*pro*	Ā-trace

Figure 4.3 Types of referential elements

But there appears to be one empty category omitted from this table, namely PRO, which does not fit neatly into any of the boxes. On the one hand, this empty category is a little like an anaphor in that it often must have an antecedent, in cases of obligatory control. But, unlike the binding of an anaphor, the control of PRO is often restricted to a particular controller, whether a subject or an object depending on the control verb. Furthermore, in cases of arbitrary control, PRO has no controller and so appears to be unbound. In the cases where PRO is bound, the question of whether it is bound within its governing category is also difficult to answer. This is because PRO always sits in ungoverned positions. This part of Control Theory is often called the **PRO Theorem**. As PRO is ungoverned, its governing category cannot be determined since this is defined by the presence of a governor. Thus PRO appears never to have a governing category. This observation allows us to explain the PRO Theorem in terms of Binding Theory. As we have discovered, pronominals and anaphors have complementary binding properties – one must be locally bound while the other cannot be. Nothing can be both an anaphor and a pronominal at the same time as such an element would be a contradiction: something that must simultaneously be bound and not bound within its governing category! Yet there is one way out of the contradiction that provides a possible analysis for PRO. An element would be able to conform to both Principles A and B of the Binding Theory, if it could do so vacuously. Two contradictory regulations can be upheld if they are always inapplicable. For example, imagine a religion which decreed both that hats must be worn in church and that hats were not permitted in church. Followers of this religion could uphold both decrees simply by never going to church! In the same way, a linguistic element can satisfy the requirement that it must be bound in its governing category and that it must be free in its governing category if it never *has* a governing category. Thus, if PRO is categorized as both an anaphor and a pronominal, we can explain its non-appearance in a governed position. Were PRO ever to be governed, it would automatically have a governing category and therefore be forced to conform to contradictory requirements. 'Since PRO is a pronominal anaphor, it is subject to Principles A and B of the binding theory, from which it follows that PRO lacks a governing category and is therefore ungoverned' (Chomsky, 1982, p. 21).

This treatment of PRO allows us to see the elements to which Binding Theory applies in a simpler light. It seems that there are things which are anaphors, things which are pronominals, things which are both anaphors and pronominals, and things which are neither anaphors nor pronominals (r-expressions). This system requires just two binary features [± anaphor], [± pronominal]. The table below demonstrates how this works out:

	+pronominal	–pronominal
+anaphor	---- PRO	reflexive A-trace
–anaphor	personal pronoun *pro*	proper noun Ā-trace

Figure 4.4 Elements of Binding Theory

Most of the cells of figure 4.4 are filled with both an overt and a covert example, the one exception being the pronominal anaphor, for which there is only a covert exemplar: PRO. However, it is fairly obvious why there is no overt exemplar of this type of DP. All overt DPs must be assigned Case, according to the Case Filter, and Case is assigned under government. Thus all overt DPs must be governed and so have governing categories. It follows that there can never be an overt pronominal anaphor as such an element would always be a contradiction.

EXERCISE 4.12 The following examples all contain empty categories which have a role in their ungrammaticalities. Which of them are ungrammatical due to Binding Theory violations and which are not?

* Who$_1$ does his$_1$ mother love t$_1$?
* John$_1$ seems for t$_1$ to like Mary.
* John$_1$ was promised PRO to leave.
* John$_1$ likes pro$_1$.
* I$_1$ expect for PRO$_1$ to leave.
* Who$_1$ did you$_1$ try t to win?
* John$_1$ thinks Bill wants pro$_1$ to leave.
* It$_1$ seems t$_1$ makes sense.

4.5.2.4 *The place of Binding Theory in the Government/Binding Model: the need for more levels*

Where does Binding Theory sit in the general GB Model we have been developing? The fact that Binding Theory has an effect on movement indicates that it must apply after movement has taken place, i.e. at S-structure. Unfortunately, however, the matter is not quite as simple as this as some binding facts seem to be established before movement. Take the following sentence:

(151) [Which picture of himself$_i$]$_j$ did John$_i$ display t$_j$?

Here the wh-phrase including the anaphor has been moved to the specifier of CP at S-structure. In this situation the subject does not c-command the anaphor and therefore should not be able to bind it. Nevertheless the sentence is grammatical. Note that at D-structure, the subject does c-command the anaphor as it is part of the phrase generated in object position at this level. But we cannot assume that Binding Theory takes place at D-structure as otherwise it would not be able to interact with movement in the ways demonstrated above.

One possible solution to this problem would be to propose that Binding Theory applies at neither D- nor S-structure, but at another level of representation separate from both, at which all the relevant binding relations can be made to hold. The next section shows that there is independent empirical motivation for such an extra level of representation, as well as conceptual arguments. The inclusion of the Binding Theory into the overall GB Model will therefore be delayed until these arguments have been presented.

4.6 Beyond S-structure and the Empty Category Principle

4.6.1 That-*trace phenomena, proper government and the Empty Category Principle*

Most of the restrictions on movement seen so far seem to be concerned with the locality of movement operations, constraining how far an element can be moved. However, some phenomena, first noted in the 1970s, suggest that it is also important to take into account the position from which a movement is made, the **extraction site**, for constraining movement. For example, movement from the object position seems to be easier than from subject position:

(152) a. What$_i$ did he say (that) Mary wanted t$_i$?
b. Who$_i$ did he say (*that) t$_i$ wanted a beer?

In (152a), as usual, the complementizer introducing a finite clause is optionally present, as represented by the parentheses, when there is a wh-movement from object position. In contrast, when a movement takes place from subject position, the complementizer must be absent. This is represented in (152b) by including an asterisk within the parentheses, stating that in this case the option of having the complementizer is ruled out. The phenomenon was noted first in Chomsky and Lasnik (1977), who termed it the ***that*-trace effect**.

Nothing so far discussed explains this. The two structures are otherwise identical; their different grammatical statuses derive solely from factors affecting the extraction sites of the movements. Of course, in the extraction site of a movement sits a trace, suggesting that the conditions under consideration here concern the licensing of traces in certain positions. It is perhaps no surprise that empty categories should need to be licensed in some way given that their presence is otherwise not marked overtly. *That*-trace phenomena indicate that traces are licensed in object position fairly freely, whereas they are licensed in subject position only in the absence of a complementizer. GB Theory proposed that empty categories are licensed by a notion of **proper government**, which we will define shortly, due to a condition known as the Empty Category Principle (ECP):

(153) **Empty Category Principle**
Traces must be properly governed.

As we have seen on several occasions, the government relationship holds between a head and elements within its local sphere of influence. Proper government was seen as a more restrictive version of government, limited to a set of 'proper governors'. Given that objects appear to be more easily licensed than subjects and also given that one difference between subjects and objects is that the former are governed by a lexical head (the verb) whereas the latter are governed by a functional head (the inflection), the obvious first step is to claim that lexical heads are proper governors but functional heads are not. Consequently the trace in an object position will always be properly governed; movement from object position might be expected to be relatively free, subject to other conditions

on movements in general. However, if this were all there was to say on the matter, given that subjects are not governed by a lexical head, a trace in the subject position should never be properly governed and hence movement from this position would be impossible. But, while movement from subject position is subject to more stringent conditions, it is not impossible and so these restrictions need to be loosened slightly.

Considering the data in (152b) again, movement from the subject position is possible, and hence the trace is licensed, when there is no complementizer. A closer look at the movement shows more clearly what is happening. Due to the boundedness of movement, the first movement to take place is from the subject position of the embedded clause to the specifier of CP of that clause. From there the wh-element will move to the higher spec CP:

(154) $[_{CP}$ Who$_i$ did $[_{IP}$ he say $[_{CP}$ t$_i$ (that) $[_{IP}$ t$_i$ wanted a beer]]]]?

Given that the intermediate trace faces exactly the same conditions as any trace involved in spec CP to spec CP movement, the conditions placed on this are irrelevant in accounting for *that*-trace phenomena. Obviously, what we want to say is that the original trace in subject position is properly governed in the absence of the complementizer, but that the complementizer blocks this government when it is present. The complementizer intervenes between the original trace and the intermediate trace and it could be the intermediate trace that acts as the proper governor. If this is so, the set of proper governors includes the set of lexical heads, plus any antecedent. The upshot is that traces in object position will always be properly governed by the lexical head they are the object of, but subjects can only be properly governed by their antecedents. Government by the antecedent will be possible only if the antecedent is relatively close and the relationship with its trace is not interrupted by a closer element, such as the complementizer.

While there is much else that could be said, this basic account will serve our purposes. The next section introduces observations which at first seem to be the exact opposite to the *that*-trace effect, suggesting that movement from the subject position is sometimes easier than from object position. The attempt to unify these observations under one set of grammatical principles leads us to extend the overall model, as already hinted at in the section on binding above.

EXERCISE 4.13 The original account of the *that*-trace effect suggested by Chomsky and Lasnik (1977) was that structures in which the complementizer was immediately followed by a trace were filtered out. This is clearly more stipulative than the ECP account of the phenomena, but consider the following observations and decide which of the two approaches is the more empirically accurate:

Who does John think that these days should be conscripted?
* Who does John think that should be conscripted?
Who does John think that these days they should conscript?
Who does John think that they should conscript?

THE EMPTY CATEGORY PRINCIPLE AND PROPER GOVERNMENT

The Empty Category Principle (ECP):
Traces must be properly governed.

Proper government:
α properly governs β if and only if:

(i) α governs β and
(ii) α is lexical or an antecedent.

Gloss:
Traces are 'licensed' by being governed by either a lexical head or an antecedent. Thus, traces in object position are always licensed as they are always governed by the head they are an object of. Traces in subject position are never head governed and therefore must be antecedent governed. This gives rise to the *that*-trace effect as antecedent government is blocked by the presence of an overt complementizer.

4.6.2 *Superiority and invisible movement*

In all of the examples of wh-movement looked at so far there has been only one wh-element per clause. However, some clauses can have more than one wh-element, forming multiple wh-questions. In English multiple wh-questions, only one of the wh-elements moves to the front of the clause and all others remain in their D-structure positions:

(155) a. $[_{CP}$ Who$_i$ $[_{IP}$ t_i saw what]]?
b. $[_{CP}$ How$_i$ did $[_{IP}$ they travel where t_i]]?

Often multiple wh-questions are interpreted as asking for an answer which pairs up possible referents for the wh-elements. For example, (155a) could be answered as follows:

(156) John saw a black car draw up, Bill saw a shadowy figure enter the house and Mary saw someone running down the drive.

This shows that the unmoved wh-element is still interpreted as an interrogative operator and not like the wh-element found in echo questions, which are also unmoved:

(157) a. You saw what?
b. They travelled where?

In these type of questions the wh-elements are interpreted as slot fillers for information that was either misheard or disbelieved: something far more specific

than the interpretation of a typical interrogative operator. This is slightly puzzling as it seems that the unmoved wh-elements in multiple questions are interpreted as moved wh-elements, though they clearly do not undergo the same movement.

Another interesting thing about multiple wh-questions is that although in principle either wh-element could move, often it is only grammatical to move one of them. For example compare the following:

(158) a Who$_i$ t$_i$ saw what?
b * What$_i$ did who see t$_i$?

Note in this case, given the choice between moving the subject or the object, it is the subject that moves more easily than the object. This phenomenon is called the **superiority effect**, as it is the structurally superior (i.e. higher) wh-element that is able to move. At first sight this appears to be exactly the opposite of *that*-trace effects, where we saw that the subject is more difficult to move. From this point of view it is difficult to think how superiority might be accounted for by the ECP, as the trace of the object is always properly governed by its verb.

We thus face two puzzles: why should the unmoved wh-element be interpreted as though it were moved in multiple questions; and why should the movement of an object cause an ungrammaticality? The solution to both puzzles lies in the assumption that there is a level of structural representation beyond S-structure and D-structure.

Let us justify the need for this extra level by considering the difference between the pronunciation of an expression and its semantic interpretation. So far we have assumed two levels of representation. Clearly S-structure is more closely associated with the pronunciation of an expression than D-structure, as the phonetic string reflects the order of elements at the post-movement level. However, S-structure is *not* a phonetic representation, but a syntactic one, as it contains all sorts of elements that have no bearing on pronunciation, such as tree nodes and branches. Thus, although pronunciation is more closely related to S-structure than to D-structure, we need to suppose a further representation which reflects pronunciation even more closely. This is **Phonetic Form**, or PF for short, as introduced in chapter 1. This would fit into the grammatical model thus:

(159) D-structure
|
S-structure
|
PF

Now let us consider where semantic interpretation enters the picture. An early idea was that semantic interpretation was more associated with D-structure, as this represents in an obvious way the thematic relations which hold between a predicate and its arguments. However, not all semantic relations are represented at D-structure. For example, many binding relationships are formed after movement rather than before:

(160) John$_i$ believes [himself$_i$ to have been promoted t$_i$].

In this example, the antecedent *John* could not bind the anaphor in its D-structure position as only the subject of the non-finite clause is accessible according to the binding principles discussed in the previous section. Therefore the binding relation between *John* and *himself* can only be properly established after the movement. Similarly, the interpretation of a wh-element as an interrogative operator is established only after movement, as at D-structure the wh-elements in interrogative and echo question structures are indistinguishable. Therefore we need to assume that semantic interpretation is taken from S-structure. Thematic relationships established at D-structure are still represented at S-structure even after movement as traces keep track of D-structure positions, so this is not problematic for this assumption.

However, certain problems remain if both semantic and phonetic interpretation are associated with the same syntactic level. For example, consider the following:

(161) Every performer sang a song.

There are several ways in which this sentence can be interpreted. In one, each performer sings a different song. In this case we say that the quantifier *every* takes wide scope over the indefinite determiner, as semantically we determine the meaning of *every performer* for each of its possible referents and then determine the referent for *a song* for each performer. On the other hand, we might interpret (161) as meaning that there was a single song that every performer sang. In this case, the DP *a song* has wide scope as its reference is determined first. In English, such scope differences tend not to be overtly reflected syntactically and the same S-structure can be interpreted in different ways. We might think therefore that the difference is purely semantic, nothing to do with the structure of the sentence. But this assumption is challenged by a number of observations. The first is that in some languages these semantic differences are indeed reflected grammatically. For example the English sentence equivalent to (161) translates into two different Hungarian sentences depending on the interpretation:

(162) a. Minden előadó elénekelt egy dalt.
(every performer sang a song)
Every performer sang a song. (wide scope on *minden előadó*)
b. Egy dalt elénekelt minden előadó.
(a song sang every performer)
Every performer sang a song. (wide scope on *egy dalt*)

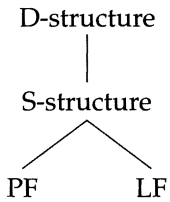
Basically, the quantifier with the widest scope comes first and hence is higher in the structure. Thus scope facts are reflected syntactically in Hungarian.

The second observation is that even in English there is an interaction between scope interpretation and certain grammatical facts. Consider the difference between the following:

(163) a. Some professor believes [every student to have failed].
b. Some professor believes [every student has failed].

Sentence (163a) is ambiguous in a similar way to (161): in one interpretation there is a single professor who thinks that all the students have failed and in the other for every student there is a professor who thinks he or she has failed. However, (163b) is not ambiguous, and specifically the reading in which *every student* has wide scope is missing. Thus, the subject of a non-finite clause can have scope over a higher subject, but a subject of a finite clause cannot. It seems therefore that scope phenomena are not solely semantic in nature and that they should be represented syntactically as well. But while in Hungarian scope is obviously represented at S-structure, it is clear that this is not true for English. Therefore we need to propose another level of representation beyond S-structure in which these facts about English can be represented. This level of representation is often called **Logical Form**, or LF, deriving from S-structure in the same way that S-structure derives from D-structure:

(164)



> It has, however, become clear that other features of semantic interpretation having to do with anaphora, scope and the like are represented not at the level of D-structure but rather at some level closer to surface structure, perhaps S-structure or a level of representation derived directly from it – a level sometimes called 'LF' to suggest 'logical form'. (Chomsky, 1986a, p. 67)

At LF, we can assume that scope facts in English are reflected in a similar way to that in Hungarian. Presumably the quantified DP with the wide scope is moved to a higher position. However, as PF reflects the syntactic arrangement at S-structure, any movement that takes place between S-structure and LF will have no effect on pronunciation and in effect such movements will be invisible.

One advantage of these assumptions is that certain differences between languages can be accounted for in a rather simple way. Take the difference between Hungarian and English with regard to scope phenomena. In Hungarian, scope facts are overtly represented and therefore the movements involved must take place between D-structure and S-structure. In English, however, the same movements are covert and so must take place between S-structure and LF. The two languages are similar in their use of the same movement processes, but they differ over where these processes take place. We can see other such differences cross-linguistically. Chinese, for example, does not appear to move its wh-elements in interrogatives, yet the grammar of Chinese is not so different from English if both languages make use of wh-movement, only English does so overtly, Chinese covertly.

We can now return to superiority phenomena. Recall that in multiple wh-questions, the unmoved wh-element is interpreted as though it moves. This can

now be handled by assuming that the apparently unmoved wh-element undergoes covert movement at LF, just like Chinese wh-elements. Presumably the covert movement is similar to the overt movement in that both move a wh-element to the CP:

(165) D-structure: $[_{CP}$ [Who saw what]]?
S-structure: $[_{CP}$ Who$_i$ [t$_i$ saw what]]?
LF: $[_{CP}$ What$_j$ who$_i$ [t$_i$ saw t$_j$]]?

This also allows superiority to be explained in terms of the ECP, but with respect to the covert movement, not the overt one. Consider the examples in (158) again, along with their respective LFs:

(166) a. [Who$_i$ [t$_i$ saw what]]? – $[_{CP}$ What$_j$ who$_i$ [t$_i$ saw t$_j$]]?
b. * [What$_j$ did [who see t$_j$]]? – $[_{CP}$ Who$_i$ what$_j$ [t$_i$ saw t$_j$]]?

In (166a), both traces are properly governed at LF: the subject's trace is properly governed by its antecedent, as it is at S-structure, and the object's trace is properly governed by the verb. However, in (166b) while the trace of the object is properly governed, as always, the trace of the subject is separated from its antecedent by the object wh-element, which moved to the CP first, at S-structure. In other words, with superiority effects, it is the covert movement of the subject that causes the problem, not the overt movement of the object, and hence this phenomenon is not unlike the *that*-trace effect.

It is clear from the above discussion that the ECP applies to LF representations, not at D-structure or S-structure. The model of the grammar can then be extended as follows:

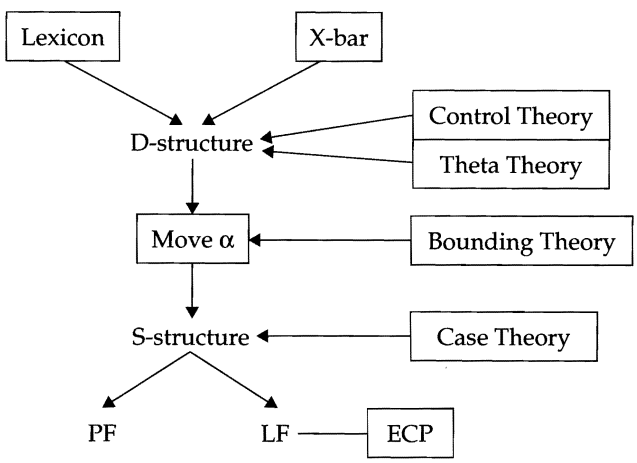


Figure 4.5 Government/Binding Theory (Logical Form and Phonetic Form added)

SUPERIORITY

Who saw what?
*What did who see?

The explanation of this is provided by the assumption that the *in situ* wh-element undergoes a covert movement at LF to the place of the other wh-element for interpretative reasons. When the subject moves first, its trace will be properly governed by its antecedent. A subsequently moved object will leave behind a trace that is properly governed by the verb:

$$\text{what}_2\ \text{who}_1\ t_1\ \text{saw}\ t_2$$

However, when the object is moved first, although its trace is properly governed by the verb, the subsequently moved subject will be unable to govern its trace properly, as the moved object will be in the way:

$$\text{who}_1\ \text{what}_2\ t_1\ \text{saw}\ t_2$$

4.6.3 *Binding Theory revisited*

The discussion of Binding Theory was left without reaching a conclusion about the level at which binding applies. There are strong conflicting arguments for it applying both before and after movement. The introduction of LF yields yet another possibility that it applies at this level. Indeed this solves a number of problems. Consider the puzzling data once more:

(167) a. [Which picture of himself]$_i$ did John display t_i?
b. [The men]$_i$ seem to each other t_i to be smart.

In (167a) the anaphor is properly bound in its D-structure position, suggesting that binding principles apply before movement takes place. However, in sentence (167b) the binding of the reciprocal anaphor *each other* is only achieved after the subject of the lower clause is raised into the higher clause, suggesting that Binding Theory applies after movement. These two apparently contradictory observations can be reconciled if Binding Theory applies at LF and if a further covert movement re-establishes the proper relationships in the case of (167a). This movement must in effect reverse the wh-movement by placing the structure containing the anaphor back into its D-structure position. Considering example (167a) more closely we can see that the overtly moved wh-phrase consists of a wh-determiner *which* and an NP containing the rest of the phrase. It is possible that the semantic interpretation of this sentence is served if only the determiner is in the specifier of CP but that the NP is moved along with the DP for syntactic

reasons holding at S-structure. In this case, we can envisage a movement which 'reconstructs' the non-wh part of the DP (i.e. the NP) back into its original position, creating the following LF:

(168) [Which t_i] did John display [picture of himself]$_i$?

If Binding Theory applies to this structure, the sentence will be correctly predicted to be grammatical as the antecedent of the anaphor binds it within the local governing category.

Although this is probably the simplest version of the **reconstruction** approach to this problem, it does leave us with a trace inside the wh-DP that is not properly governed and so in violation of the ECP. In order to get round this problem some more complicated assumptions would have to be made. These will not be pursued here as it will turn out in subsequent chapters that the problem does not arise under current assumptions.

LOGICAL FORM

Logical Form (LF) is another level of syntactic representation like D-structure and S-structure. It is produced from S-structure in the same way that S-structure is produced from D-structure: via movements. However, as Phonetic Form (PF) is taken from S-structure, the movements which form LF are not phonetically realized and have purely semantic repercussions. Languages may differ as to where certain movements take place: at S-structure or at LF. If the former, then the movement will be visible – e.g. wh-movement in English or quantifier movement in Hungarian – and if the latter, then the movement will be invisible – e.g. wh-movement in Chinese or quantifier movement in English. LF is also the level of representation at which reconstruction movements take place, putting back material into a position where it is more readily semantically interpreted from a position it was forced to move to by S-structure requirements.

Discussion topics

1 Movement is seen in GB Theory as a relationship between D- and S-structure. Do you think this works satisfactorily for apparent variants like 'I gave him a book', 'I gave a book to him' and (in the speech of one of the authors) 'I gave it him' versus 'I gave him it' or 'Whom did you give it to?' and 'To whom did you give it?'?

2 Given that one goal of the grammar is to achieve explanatory adequacy, what does the GB Theory of movement imply for children's acquisition of language?

3 Move α is a very liberal principle essentially allowing anything to move anywhere. Take a simple sentence such as *the choir sang a song* and attempt to move any element into any other position. Do we know why not all of these movements produce grammatical sentences?

4 The Case Filter states that all DPs must have Case. Is it possible for one DP to

have more than one Case? Can a DP, for instance, move from one Case position to another?

5 Sometimes we use reflexive pronouns without apparent antecedents, as in:

> Only John and myself knew the answer.
> What's a good kid like yourself doing in a place like this?
> Address the envelope to yourself.
> Lies about yourself can be very upsetting.

Are all of these uses of reflexives problematic for Binding Theory?

6 Some languages assign inherent Case (dative or genitive) to the subjects of certain Verbs and in these cases the object shows up in the nominative. This has been claimed to be another example of Burzio's generalization (p. 161). Can you see why?

7 Suppose that all quantifiers move to the front of the clause by LF. Would the ECP predict possible scope ambiguities between two quantifiers by having either move higher than the other?

8 In Government/Binding Theory, how many principles prevent an element from moving too far? Does the fact that many principles of the grammar appear to have similar functions detract from the elegance of theory?

9 Towards the end of the 1980s Chomsky suggested that certain movements were made only as a last resort. Thus if something didn't have to move it wouldn't. For example, an inflection will not move to a verb if the verb can move to the inflection: English auxiliaries, which are not stuck inside the VP, always move to I. Does this kind of observation fit well with the general conception of Move α?

5 Chomskyan Approaches to Language Acquisition

This chapter introduces the approach to language acquisition that springs out of Chomsky's Principles and Parameters (P&P) Theory, developing some of the ideas about acquisition seen in chapter 2 and relying chiefly on the syntax described so far. We will concentrate on first language (L1) acquisition here, leaving second language acquisition (SLA) to the next chapter. Starting with some general concepts of acquisition, we then move on to a particular issues in parameter-setting theory. The chapter does not give a comprehensive account of recent research, for which the reader is referred to Guasti (2002), but aims to illustrate the themes in Chomskyan-style research. Mostly the discussion uses the terms of Government/Binding (GB) Theory since this is still the most used framework for acquisition research.

5.1 The physical basis for Universal Grammar

Acquisition of language is, to Chomsky, learning in a peculiar sense: it is not acquisition of information from outside the organism, as we acquire, say, facts about geography; it is not like learning to ride a bicycle, where practice develops and adapts existing skills. Instead it is internal development in response to vital, but comparatively trivial, experience from outside. To make an analogy, a seed is planted in the ground, grows and eventually flowers; the growth would not take place without the environment; it needs water, minerals and sunshine; but the entire possibility of the plant is inherent in the seed; the environment only dictates the extent to which its inherited potentialities are realized. Knowledge of language needs experience to mature – without it nothing would happen – but the entire potential is there from the start. Chomsky argues that language acquisition is more akin to growing than to learning; it is the maturing of the mind according to a preset biological clock. 'In certain fundamental respects we do not really learn language; rather grammar grows in the mind' (Chomsky, 1980a, p. 134). Language is part of the human inheritance; it is in our genes. As Lenneberg (1967) pointed out, to become a speaker of a human language does not require a

particular size of brain; it does not require a particular type of interaction with adults. The requirements for learning a human language are to be a human and to have the minimal exposure to language evidence necessary to trigger the various parameters of Universal Grammar (UG).

The physical basis of UG means that it is part of the human genetic inheritance, a part of biology rather than psychology; 'universal grammar is part of the genotype specifying one aspect of the initial state of the human mind and brain' (Chomsky, 1980a, p. 82). As with other inherited attributes, this does not rule out variation between individuals. Most introductory linguistics books assert that all human children can learn all human languages, a difficult statement to substantiate even if intuitively correct; empirical evidence would presumably have to come from children from one type of human being brought up by parents from another. Though it is rather indirect evidence, Korean children adopted in France under the age of 5 grow up with linguistic competences indistinguishable from those of children born of French parents (Pallier et al., 2003). If language is indeed due to common genetic inheritance, some individual variation might be expected; all human beings have eyes but some are brown, some blue, some green. At the moment the common features of UG in all human minds need to be established before variations between individuals can be investigated. Controversial work by Gopnik and Crago (1991) has, however, claimed to find a particular language deficiency widely spread among the members of one family, suggesting a common genetic origin. For the moment it is how the human eye works that is of basic importance, not variations of colour in the iris. 'The main topic is the uniformity of normal development. It could ultimately be an interesting question whether there is genetic variation that shows up in language somewhere' (Chomsky, 1982, p. 25). While the bulk of UG seems common to all human beings, this assumption is not at present based on proof that, say, 'English' native speakers of Japanese have identical competences to 'Japanese' speakers of Japanese.

The language organ is also held to be the property only of human brains, as we saw from Lenneberg (1967): people speak; dogs do not. Various attempts have been made to refute Chomsky's claim that the language faculty is species-specific. It is clear that animals are aware of more subtle aspects of language than people had previously accepted; Gentner et al. (2006) for example claim to show that starlings can deal with recursive syntax; Toro et al. (2005) showed that rats were able to distinguish Dutch from Japanese intonation. Most humans, however, could tell the difference between the sound of a Rolls Royce and a Trabant, but that does not mean that they are able to speak 'engine'!

The usual approach is to teach another species a simplified form of language, substituting some other means of expression for the vocal apparatus, such as gestures or visual signs. This approach was adopted in a series of experiments with apes in the 1970s and 1980s. The general assumption was that the reason why apes could not talk was not so much their lack of mental capacity as their physical difficulties in producing the sounds of human speech, in particular the different structure of their larynx. Several projects substituted another medium for spoken language, assuming that this would open the door for the apes to acquire language.

Gardner and Gardner (1971) brought up a chimpanzee called Washoe in ways similar to a human child encountering American Sign Language (ASL), the gesture language used by the deaf in the USA; this was taught primarily by 'modelling', that is to say, the human guided the chimp's hands to make the appropriate sign. Considerable success appeared to be achieved by this means. Washoe for instance produced the subject *you* sign before the object *me* sign 90 per cent of the time these were combined, i.e. she seemed to know the correct subject–object order. Washoe too would react differently to the word orders in *baby mine* and *mine baby*.

Rumbaugh (1977) taught an invented language to a 2-year-old chimp called Lana on the keyboard of a computer; this 'Yerkish' language consisted of abstract symbols, such as ⇔ for 'coconut', which were displayed on a computer monitor, her only means of communication. She was guided through a careful sequence of increasingly complex 'sentences' by reinforcement with sweets etc. She too appeared very successful; the first time she was shown an orange, she produced the sequence *Tim give apple which-is orange*.

Patterson (1981) taught a 1-year-old gorilla called Koko a version of ASL. By the age of 5½, she had mastered 246 signs of ASL, such as 'alligator', 'cake', 'small' and 'pour'. Again several combinations were produced by Koko that she could not have encountered, for example *white tiger* for a toy zebra, *bottle match* for a cigarette lighter, and *cookie rock* for a stale bun. Superficially these studies seem to show that the apes are using language to communicate, that they can make use of order to signal meaning, and that they are capable of creating new utterances when the situation demands it.

But there are many reasons why this evidence cannot be held to prove the acquisition of language by apes, clearly presented in Wallman (1992). Much of the research has problems of methodology and interpretation; it is all very well to report that Koko produced *cookie rock*; what she actually did was produce two gestures interpreted by the human observers as *cookie* and *rock*; what they meant to her we do not know. It seems at best language-like behaviour that, even compared with a human child, does not begin to tap the complexity of human language. Chomsky's ironic comment compares how well human beings fly with how well apes speak; perhaps 'the distinction between jumping and flying is arbitrary, a matter of degree; people can really fly, just like birds, only not so well' (Chomsky, 1976, p. 41).

One crucial objection is that the languages involved do not contain anything resembling the principles of UG: whatever the ape has acquired, it is not core grammar. The ape's knowledge might be peripheral or it might be functional knowledge of how to achieve things through gestures or signs, but it is not language knowledge in the UG sense. Chomsky's main objection is that the learning described does not resemble language acquisition in children because it is taught rather than 'picked up'. Children learn language from positive evidence rather than from reinforcement by their parents. The apes, however, are all *taught* language; in the wild, apes do not develop language for themselves. The child 'does not choose to learn, and cannot fail to learn under normal conditions' (Chomsky, 1976, p. 71). The role of the environment to the child is triggering; it sets things off rather than providing precise, controlled instruction. Even if the attempt to

teach language to apes were successful, it would not prove how animals could *acquire* language. To sum up: 'the interesting investigations of the capacity of the higher apes to acquire symbolic systems seem to me to support the traditional belief that even the most rudimentary properties of language lie well beyond the capacities of an otherwise intelligent ape' (Chomsky, 1980a, p. 239). At best the experiments with apes demonstrate 'that apes raised by human beings in a human-like environment . . . can develop some human-like skills that they do not develop in their natural habitat' (Tomasello, 1999, p. 35).

Let us go back to the general issue of where language comes from. Each individual child is on their own in acquiring language; in a sense each of them has to construct language for themselves from the meagre data available to them. The fact that most of them succeed and end up with essentially the same grammar despite vast variations in situation and language input must demonstrate the sheer propensity of their minds to acquire language.

Some children in fact have to go further, if, say they have parents who speak a pidgin language that is not spoken by anybody as an L1 and does not have the characteristics of a native language, e.g. the pidgin English that developed amongst black slaves in British colonies in the West Indies. When their children learn to speak, they somehow turn this input into a creole with all the characteristics of a native language. In a sense they are not so much creating an individual language in their minds as creating a new language altogether. Derek Bickerton (1984) believes this is the moment when the innate linguistic powers of the mind come into operation through what he terms the 'bioprogram'. The difference from the UG position is that, while UG is used by every human child to acquire language, the bioprogram is used only by children who get input that is not a fully fledged language.

Perhaps the first time that such a language has been captured in the act of creation is the fascinating case of Nicaraguan Sign Language (Senghas et al., 2004). When deaf children were brought together into a school in Nicaragua in the 1970s, it was the first time that a community of deaf people existed in the country who needed to communicate with each other. Their previous language experience had been confined to lip-reading and spontaneous home-signs. Nevertheless the first generation of these children managed to create a sign language essentially as a pidgin that was nobody's L1. The next generation of children arriving at the school then created a creole language out of the pidgin they encountered, which was indeed their L1. The grammatical system is being created by the younger children, who use it with more fluency than the older children. As Senghas and Coppola (2001, p. 326) say, 'The mark left by this generation of deaf Nicaraguans is an entirely new language, one that did not exist when they were born.' Somehow human children can create languages even when the input they hear is not a full language. Senghas et al. (2004) show how the younger children are encoding complex motions more linguistically as a sequence of signs than the older children who are closer to the natural gesturing system of speakers. It will be interesting to see whether the Nicaraguan Sign Language corresponds not just to the overall idea of the built-in potential of the human mind to create new languages, as it clearly does, but also to the specific predictions of the UG Model in terms of principles and parameters.

5.2 A language learning model

It is useful to break down the language learning situation into components to see the role and importance of various aspects of the problem facing the language learner. A common view identifies three main components of language learning: the **search space** which limits the possibilities of what there is to be learned, the **evidence** on which learning is to proceed and the **learning strategies** applied by the learner. Although in formal language learning theory much attention has been paid to the issue of the search space, investigating properties of formal languages and their grammars which might affect the learning situation, the assumption of UG greatly simplifies this aspect of learning. The search space now consists of the set of parameters and their possible values: the child has to learn which parameter is set to which value in the target (adult) language. As there are a finite number of parameters each with a finite number of possible values, the search space itself is finite; this fact alone eliminates many potential learning problems, though not all of them. The next sections concentrate on the issues of evidence and learning strategies.

5.2.1 The evidence available to the first language learner

A constant theme in Chomsky's writings is the nature of the evidence available to the child, taking up arguments suggested by Gold (1967), Baker (1979) and others. Children have to acquire a language from the evidence they encounter. Without any evidence at all, they will acquire nothing; with evidence, they will acquire Chinese or Arabic or any human language they encounter, including sign languages. The 'logical problem of language acquisition' hinges upon the types of evidence they meet and the uses they put it to.

This can be illustrated by looking at how one might learn to play games such as chess or snooker. Let us take snooker as an example – the reader who does not know the game will find the example even more convincing! After years of watching snooker on television, John Smith knows from observation some of the sequences of colours in which balls are hit by the players, for example every other ball hit into a pocket has to be red and the last ball to be pocketed has to be black. If he started to play snooker tomorrow, he could copy the sequences he has already observed. But he would not be able to tell if a new sequence, say two red balls in succession, was illegal, or simply one he had by chance not encountered before. He would have no idea what sequences are actually impossible because, given the standard of competition play, he has only seen sequences in which the rules are obeyed rather than those in which they are broken. While he might pass as a snooker player for a few minutes, an adequate knowledge of snooker involves knowing what *not* to do as well as what to do. To learn snooker properly, he would need to get some other type of evidence. One possibility is to see players playing a foul shot, a rare event in professional television snooker. Furthermore, to recognize a sequence as a mistake, something must indicate that it is wrong, such as a penalty from the referee (i.e. a 'snooker') or the hissing of

the crowd or a remark from the commentator; otherwise it would simply appear to be another permissible sequence to add to his stock. A further possibility is to learn from the mistakes he makes while actually playing; he hits a black ball followed by a blue and sees if his opponent tells him it is wrong. Or he might deduce from the fact that he has never seen a particular colour sequence, say blue followed by black, that it is illegal; this would not help him to distinguish sequences that are impossible from those that are rare or unlikely. Finally he might buy a guide to snooker and read up on the actual rules given there; this solution is no use if he is unable to read or cannot grasp the type of information given in the rule, as would be the case for the child acquiring the L1. Overall it would be impossible for him to learn the rule of colour sequences from just watching snooker games in progress. At best he will have a fair idea of them from the sequences he has observed, but does not know how representative these are of all the games he has not observed.

This analogy illustrates the general properties of the evidence that are necessary for acquisition. On the one hand there is **positive evidence** of actually occurring sequences, i.e. sentences of the language. On the other hand is **negative evidence**, such as explanations, corrections of wrong sequences or ungrammatical sentences, that shows what may *not* be done. A knowledge of correct snooker sequences appears to be unlearnable only from positive evidence from games that are actually witnessed, and requires additional evidence from impossible sequences, corrections, ability to read the rule-book, ability to comprehend abstract explanation, and so on. But these possibilities are not available in L1 acquisition; except for positive evidence, the other forms detailed above – correction, explanation etc. – are seldom encountered by the child. The foundation of UG accounts of language acquisition is that evidence other than positive evidence by and large cannot play a critical role; the child must learn primarily from positive examples of what people *do* actually say rather than negative examples of what they *don't* say. Again, this comes back to the poverty-of-the-stimulus argument introduced in chapter 2: to acquire language knowledge from experience, the mind needs access to evidence other than actual sentences; since this is not available, the knowledge cannot be acquired but must already be there.

Chomsky (1981a, pp. 8–9) recognizes three logically possible types of evidence for acquisition. First comes 'positive evidence (SVO order, fixing a parameter of core grammar; irregular verbs, adding a marked periphery)'. The occurrence of particular sentences in the speech children hear tells them which sort of language they are encountering and so how to set the parameters; hearing sentences such as:

(1) The hunter chopped off the wolf's tail.

they discover that English is head-first in that the Verb head *chopped* comes before the verb complement *the wolf's tail*. Hearing sentences such as:

(2) Ojiisan wa momo o kirimashita.
(the grandad the peach cut)
The grandad cut the peach.

they discover Japanese is head-last in that the verb head *kirimashita* comes after the object *momo o*. Positive evidence can set a parameter to a particular value.

Chomsky's second type of evidence is 'direct negative evidence (corrections by the speech community)'. The child might conceivably say:

(3) Man the old.

and a parent correct:

(4) No dear, in English we say 'The old man'.

In other words the child's mistake of putting the noun before the article and adjective is corrected overtly.

The third of Chomsky's evidence types is 'indirect negative evidence'; the fact that certain forms do *not* occur in the sentences the children hear may suffice to set a parameter. An English child is unlikely to hear many subjectless declarative sentences:

(5) *Speaks.

or subject–Verb inversion:

(6) *Speaks he.

save for performance mistakes. At some point the cumulative effect of this lack might push English children into deciding that English is a non-pro-drop language. Chomsky claims that 'There is good reason to believe that direct negative evidence is not necessary for language acquisition, but indirect negative evidence may be relevant' (Chomsky, 1981a, p. 9). This division into three types of evidence will be used in later discussion, with some qualifications.

The argument partly depends on two requirements which, though Chomsky does not name them explicitly himself, can be called **occurrence** and **uniformity**. It is not enough to show that some aspects of the environment logically could help the child; we must show that this *does* occur. While it is at least conceivable that parental explanation of a grammatical principle, such as the Case Filter, might be highly useful to the child, it is inconceivable that it actually occurs. If a model of acquisition depends crucially on children hearing a particular structure or on their being corrected by their parents, it is necessary to show that this does actually happen; to meet the occurrence requirement, speculations about the evidence that children might encounter need support from observations of what they *do* encounter.

All children with very few exceptions learn language. Suppose a learning theory suggests that acquisition depends upon the provision of particular types of evidence, say certain types of feedback from their parents, and that observations of children confirm that these do occur. This explanation would still be inadequate if a single child were found who had acquired language without this type of evidence. Some children are corrected by their parents, some are not, yet

all acquire language, so acquisition cannot crucially depend upon correction. Since language knowledge is common to all, the uniformity requirement stipulates that a model of acquisition must only involve properties of the situation known to affect all children. Uniformity is a stronger form of occurrence; it is not enough to show that a certain type of evidence is available to one child; it must be available to all children. Children are capable of acquiring their L1 despite wide differences in their situations within a single culture and across different cultures. Baby-talk words such as *puff-puff* or *bow-wow* are used by some parents in England and shunned by others; some children are told *Open the window*, some asked *Could you open the window?*; yet, widely as children differ in the extent to which they are able to use language, they nevertheless attain the same grammatical competence. So long as the environment contains a certain amount of language input, it appears not to be critical for the acquisition of grammatical competence exactly what this sample consists of; any human child learns any human language, whatever the situation they encounter. As Gleitman (1984, p. 556) succinctly puts it: 'Under widely varying environmental circumstances, learning different languages, under different conditions of culture and child rearing, and with different motivations and talents, all non-pathological children acquire their native tongue at a high level of proficiency within a narrow developmental frame.'

One might add a third, weaker, requirement. Given that children have a particular type of evidence available to them, we must still demonstrate that they take advantage of the opportunity afforded to them; this can be called the **take-up requirement**. As any teacher knows, an explanation can be offered to a student several times without them showing any signs of taking it in.

What the child hears – the input – is nevertheless vital to the P&P Theory. What children need is raw linguistic data to get their teeth into. Without hearing any words of the language, they would have nothing to say; without hearing sentences, they would not be able to set the parameters appropriately for the language they are acquiring.

The poverty-of-the-stimulus argument, which maintains that the evidence is inherently too impoverished for the child to be able to acquire UG principles, must be firmly distinguished from an early claim made by Chomsky (Chomsky, 1980b), namely the **degeneracy of the data**. Language acquisition is made more difficult by the fact that children encounter performance and so the sentences they hear may contain mistakes, slips of the tongue, English sentences with null subjects, and so on. Part of the positive evidence is therefore misleading and has to be filtered out by the learner. A model of language learning cannot be predicated on children hearing only grammatical sentences; it has to be able to tolerate a certain amount of ungrammaticality (Braine, 1971).

However, the claim that the input data for the child are degenerate has had to be modified in the light of the research into speech addressed to children, which showed that, on the contrary, such input was highly regular. Newport (1977) found that only 1 out of 1,500 utterances addressed to children was ungrammatical. Such regularity, however, applies only to the language addressed directly to the child; children doubtless also hear adult performance addressed to fellow adults, with the usual quota of deviancies. The main argument concentrates on the poverty of language addressed to children – the fact that it does not contain the right kind

of syntactic evidence – rather than on the degeneracy of the data – the fact that it is not always completely well formed – since the data are arguably not as degenerate as was earlier thought.

Many claims and counterclaims about the speech addressed to children have indeed been made in the language acquisition field at large. Bruner (1983), for example, postulates a Language Acquisition Support System (LASS) in adults' minds that enables them unerringly to provide the appropriate environment for their children; 'Language is not encountered willy-nilly by the child; it is shaped to make communicative interaction effective – fine-tuned' (Bruner, 1983, p. 39). Even if such fine-tuning, or indeed any other adaptive linguistic behaviour by the parent, could teach the principles of UG, it would still have to meet the uniformity requirement that all parents use it, which seems unlikely.

Input itself is nonetheless vital to the UG Model, as we saw above. Without hearing an example of *see*, children cannot begin to classify it as a Verb. Without hearing English sentences such as:

(7) John sees Mary.

they cannot build up the lexical entry for *see* with the object NP that has to follow it. Without hearing sentences such as:

(8) John sees himself.

they cannot discover that *himself* is an anaphor. Without hearing:

(9) It's unlikely.

they may not be able to discover that English is a non-pro-drop language, making use of pleonastic subjects rather than a phonologically empty *pro*. Language evidence is crucial to the process of acquisition. For children to be able to create linguistic competence in their minds, they need to hear a range of sentences from adults. But the speech they hear does not need to be specially adapted to them in any way.

Or does it? James Morgan (1986) has claimed that children will not be able to tell how to set a parameter from the straightforward evidence they hear in some cases. Let us take an idealized example to demonstrate this. Suppose a child hears:

(10) The dog bites the cat.

How would the child know if this were a language with an SVO (Subject–Verb–Object) order, like English, meaning 'the dog is biting the cat', or one with an OVS (Object–Verb–Subject) order, meaning 'the cat is biting the dog', rare as the latter may be in actual fact? Only if the input somehow showed this by 'bracketing' together the VP, i.e.:

(11) The dog [bites the cat].

Morgan (1986) therefore claimed that phonological clues such as pauses and intonation show the child where the syntactic boundaries come. He found in an experiment that speech addressed to children indeed has clearer marking of phrase boundaries through pauses and vowel lengthening than speech to adults. He argues that children need 'bracketed input', that is to say sentences with clear signs of particular phrase boundaries, and that parents actually provide such input.

Much of Morgan's work is within the theory of learnability, a mathematical approach to language acquisition that has had a symbiotic connection to the UG Theory. Formal learnability theory, starting with the work of Gold (1967), is concerned with what is logically necessary for language learning. Wexler and Culicover (1980) measured the input to the learner in 'degrees' going from degree-0 (no embedded clauses) through degree-1 (1 embedded clause) upwards; their argument is that children need to encounter degree-2 sentences to acquire human language, i.e. sentences containing two embedded clauses. Morgan (1986) suggested that bracketed input reduced the degrees of embedding necessary to one, i.e. degree-1; children could learn a human language if they heard sentences with at most one embedded clause. Lightfoot (1989, 1991) attempted to reduce this still further within the P&P Theory. He claimed that all the information that the child needs to learn any kind of sentence is present in sentences with no embedded clauses, i.e. degree-0. This raises problems about the properties of embedded clauses that do not seem to be directly represented in the structure of the main clause of a sentence. Movement restrictions from one clause to another are one example; Lightfoot (1991) claims that these can be triggered by information as to whether movement within the simple clause is possible, such as that in the French:

(12) Combien as-tu vu de personnes?
(how-many have-you seen of people)
How many people have you seen?

Another problem is Binding Theory, where the behaviour of anaphors and pronominals in sentences such as:

(13) Helen said she met her.

seems unlearnable with only a single clause to play with. Lightfoot (1991) modifies his degree-0 claim by saying that it is not the single clause that is needed so much as the single binding domain, as described earlier; 'a child with access only to material from unembedded binding Domains has access to certain well-defined elements of an embedded clause' (Lightfoot, 1991, p. 31). He claims that this degree-0 property is an ancillary to UG; it reflects learning strategies, not the language faculty itself. Most of Lightfoot's arguments, however, depend upon historical change in language, where the degree-0 argument can be applied; for example the change from Old English OV order to Modern English VO order was facilitated by the increasing possibility of allowing particles to separate from verbs. (Denison (1994), however, doubts whether there is sufficient evidence to be able to label Old English conclusively as SOV.)

EXERCISE 5.1

Here are some examples of parents' speech to children taken from the Manchester and Bristol transcripts in the CHILDES database (http://childes.psy.cmu.edu/). Decide what these sentences might be evidence for and whether they provide positive or negative evidence:

1 There he is, look.
2 Is that a table?
3 Don't be silly.
4 You don't want to put all your toys in the potty.
5 Don't you say pick it up please?
6 If you don't speak to Simon properly then he won't.
7 Are you making a big mess?
8 We don't want to get pen on Mickey Mouse.
9 Don't know what you're talking about.
10 I don't think it is.

5.2.2 *Learning strategies*

Let us now turn our attention to the alternative strategies that children could adopt for learning language. This brings together various points against non-UG positions, derived from Chomskyan thinking in general rather than a specific source. To give concreteness, wherever possible the discussion uses examples of the language of young children and their parents taken from the Bristol transcripts in the CHILDES database (http://childes.psy.cmu.edu/).

5.2.2.1 *Imitation*

Children might learn by imitating the behaviour of those around them. Children brought up among Italian-speaking people speak Italian, not English; their knowledge of language reflects their experience. If this counts as imitation, the child learns by imitating. However, often the term 'imitation' has been applied more specifically to speech exchanges in which children repeat the speech of adults (e.g. Clark and Clark, 1977, p. 334), as in:

(14) TV: It's Tuesday.
Child: It's Tuesday.

The basis of this acquisition model is that children parrot what is said to them.

In terms of the present discussion, children can only imitate what they actually hear. Imitation is totally based on positive evidence, with all its deficiencies; principles such as the Case Filter could not be learnt by imitation because what they hear is sentences, not grammatical principles. Furthermore speakers can produce new sentences that have never been produced before and can understand sentences that they have never met before: language knowledge is creative, as we saw in chapter 1. However often children imitated the speech of others, they would be unable to produce new things they had never heard before. Imitation

in this sense cannot account for the vital creative aspect of language use. The defence to this charge is that children in some way generalize to new circumstances they have not met before, though of course this is not, strictly speaking, imitation; Chomsky insists that generalization is not an adequate explanation because it conceals 'a vast and unacknowledged contribution . . . which in fact includes just about everything of interest in this process' (Chomsky, 1959, p. 58). In addition what native speakers know, and what children learn, is an I-language. What children hear, however, is externalized language (E-language) sentences, not internalized language (I-language). They could not acquire competence from such evidence, as I-language cannot be acquired from examples of E-language alone.

Moreover, imitation alone is not enough even to begin to learn a language. The daughter of one of the authors returned home from nursery school and announced that she had learned to count in German through learning a poem. She proceeded to demonstrate her knowledge by counting out on her fingers the first three numerals: *ein, swei, policei!* ('one, two, police!' – the German nursery rhyme is similar to the English 'one, two, buckle my shoe!' where every third phrase rhymes with the preceding number). Clearly one does not gain much understanding through merely repeating the words of others.

Imitation turns out to be rare in transcripts of interactions of English-speaking parents and children but is common for example among the Kahuli of Papua New Guinea (Ochs and Schieffelin, 1984). The occurrence requirement seems culturally determined; that is to say, imitation does not occur in all the situations children are raised in, and so the uniformity requirement that all children must encounter it is not met. Direct imitation in the form of repetition of adult sentences is unlikely to lead to acquisition.

However, to some extent this may be a straw-man argument; staunch supporters of direct imitation are thin on the ground. Deferred imitation in which the child repeats an adult remark some time later may have a greater frequency in the child's speech than direct imitation and hence meet the occurrence requirement, though it may be hard to detect in the child's speech. Nor is frequency in itself necessarily important; even if the child only imitated once a day, it might still be the key to acquisition. Nevertheless it is still true that children cannot imitate what they do not hear: if they never hear relevant sentences, they will never be able to imitate them.

5.2.2.2 Direct teaching

Although direct teaching is not a strategy adopted by the learner, it has a place in the present discussion as many parents assume they teach their children language. Due to the 'take-up' requirement mentioned above, in order for teaching to be useful in the acquisition process, language learners must be receptive to it; hence they would not be merely passive participants if this were how language learning took place. There are two sub-parts to this: **explanation** and **correction**.

Hypothetically, the binding principles for *them* and *themselves*, for example, might be taught to the child through parents explaining that there are two classes of words, anaphors and pronominals, which behave in different ways. Explanatory evidence in principle could compensate for the inadequacy of positive evidence. The minds of students learning foreign languages or computing languages are

often presumed to work in this way by their teachers. But it is totally implausible for L1 acquisition. First, such conscious knowledge of language, essentially similar to the linguist's knowledge, is different from unconscious competence and must be the property of some faculty of the mind other than language. It is also doubtful whether young children could acquire such abstract and complex conscious knowledge: a child that is old enough to understand the explanation is hardly in need of it.

Second, the occurrence requirement requires a search for such explanations in the speech of parents to children. Not only have few instances of syntactic explanation by parents been found in transcripts, but also most parents do not possess sufficient conscious knowledge of abstract UG principles to be able to give explanations of them – how many could explain the Case Filter, for example? Though made in a second language (L2) learning context, Chomsky's point is equally applicable to L1 acquisition: 'one does not learn the grammatical structure of a second language through "explanation and instruction" beyond the most elementary rudiments, for the simple reason that no one has enough explicit knowledge about this structure to provide explanation and instruction' (Chomsky, 1972a, pp. 174–5).

Finally, the uniformity requirement requires all children to encounter syntactic explanation, which seems unlikely. So, while some aspects of language might well be learnt through explanation, for example politeness forms such as terms of address for different relatives *Auntie Nellie* and *please* and *thank you*, its difficulty and its rarity suggest it can hardly be the prime means of learning principles of syntax.

Adults might, however, provide negative evidence by explicitly correcting the child's malformed sentences, i.e. by reinforcement. A child might say:

(15) Book the blue.

and the parent might dutifully correct:

(16) No we don't say that. We say 'the blue book'.

Like explanation, correction could in principle compensate for deficiencies in the positive evidence; even if parents themselves don't supply negative evidence, they might react to the child's own mistakes. This view of language acquisition as a process of teaching by adult correction is one that seems to be commonly held by parents.

For correction to be feasible, the occurrence requirement has to be met by finding evidence not only that children indeed produce ungrammatical sentences but also that adults correct them. Starting with the children's speech, examples can be found such as:

(17) I broked it in half.
She be crying because her fur will get wet, wouldn't she?
What did my mummy do at you?

Since these are ungrammatical in terms of adult competence, the actual occurrence of ungrammatical sentences is clearly demonstrated. If children learn the

head parameter for the order of elements in phrases by correction, however, they must produce sentences that specifically violate the appropriate setting. The child would have to make mistakes such as the concocted example:

(18) Man old the go will.

So that the adult can point out:

(19) No. You should say 'The old man will go'.

Though the real children's sentences above are typical and familiar to every parent, sentences that go against the typical order within the phrase are hard to find; Roger Brown (1973, p. 77) for example comments that 'the child's first sentences preserve normal word order'. Put conservatively, children do not seem to make many mistakes with UG word order principles; from the first time they use a principle, they get it right.

While the occurrence requirement is met in that children do produce ungrammatical sentences, they have not been shown to produce sentences that actually violate UG – which explains why the sentence *Man old the go will* had to be concocted, not real. Chomsky asserts: 'Though children make certain kinds of errors in the course of language learning, I am sure that none make the error of forming the question "Is the dog that in the corner is hungry?" despite the slim evidence of experience and the simplicity of the structure independent rule' (Chomsky, 1971, p. 30). Though it is impossible to prove the negative point that relevant mistakes never occur, they are, to say the least, extremely hard to find. But, like any evidence from absence of particular phenomena rather than their presence, there is always the possibility that new evidence might turn up to disprove it, though it would remain difficult to explain how children could learn language on the basis of such sparse evidence.

Even if the occurrence requirement were met in the child's speech, it still has to be shown that correction occurs. A preliminary point is that, in order to correct the mistake, parents have to be able to detect it. But some mistakes are not apparent to the listener. If the child said:

(20) Peggy hurt her.

with the meaning:

(21) Peggy_i hurt her_i.

i.e. 'Peggy hurt herself', the fact that *her* was being used incorrectly could not be detected if the listener finds an alternative plausible interpretation which happens not to be the one intended, i.e. 'Peggy hurt another female person.' Berwick and Weinberg (1984, p. 170) make the same point in terms of parsing: 'If an antecedent can be found, the sentence will be grammatical, otherwise not: in both cases the sentence will be parsable.'

Let us use the term 'correction' to refer to explicit comments by the adult on the form of the child's speech. Take the following exchange between a Bristol mother and child:

(22) Child: I yeard her.
Mother: You *heard* her.
Child: Yeard her.
Mother: Not yeard. Heard.

Such examples bear witness that correction does indeed occur; the problems are *how often* it occurs and *what* is corrected. The above example was one of only five that could be found in six Bristol transcripts amounting to some three hours of recording of diverse activities. A second example was:

(23) Child: I'm calling it a flutterby.
Mother: That's wrong, isn't it?

The remaining three corrections concerned *please* and *thank you* as in:

(24) Child: Find some more.
Mother: Please. Ask him properly and he might.

In only one of these five corrections does an adult directly point out what is wrong with the syntax of the child's speech and even this example may be simply correction of a well-known baby-talk spoonerism *flutterby*: if this is typical, explicit correction of syntax is a rare phenomenon. Howe (1981) found only one occurrence of correction of well-formedness in 1,711 mothers' replies to children. Brown (1973, p. 412) comments 'in general the parents seemed to pay no attention to bad syntax, nor did they even seem to be aware of it'. When Brown and Hanlon (1970) correlated the grammaticality of the children's speech with approval or disapproval by the mother, they found that some sentences were frowned on:

(25) Child: And Walt Disney comes on Tuesday.
Mother: No he does not.

simply because they were factually untrue, while other sentences were praised:

(26) Child: Draw a boot paper.
Mother: That's right. Draw a boot on paper.

where the meaning was obvious though the grammar was wrong. Often then grammatical sentences are corrected and ungrammatical sentences are approved, as the examples above show – only one correction out of the five related to an ungrammatical sentence. Hirsh-Pasek et al. (1984) found a ratio of 3 to 1 for approval of well-formed versus ill-formed sentences and 5 to 1 for disapproval of well-formed versus ill-formed, suggesting that something other than grammaticality is involved.

Even if children get appropriate correction, the take-up requirement still needs to be met: do children actually pay any attention to correction? A famous exchange reported in McNeill (1966) is:

(27) Child: Nobody don't like me.
Mother: No, say 'nobody likes me'.
Child: Nobody don't like me.
[eight repetitions of this dialogue]
Mother: No, now listen carefully; say 'nobody likes me'.
Child: Oh! Nobody don't likes me.

First, one is struck by the sheer ineffectiveness of the mother's repetition; the take-up requirement is not met. Second, one feels slightly uncomfortable with the dialogue: why is the mother not reassuring the child that someone *does* like them? For the usual adult response to children's speech is to comment not on its grammaticality, but on what it means (Brown and Hanlon, 1970).

However, there may again be an element of tilting at windmills in the argument: even if parents seldom correct the syntactic structure of their children's sentences overtly, correction could take more subtle forms. For instance the child's sentence given above:

(28) Draw a boot paper.

was not corrected directly by the mother but was expanded:

(29) That's right. Draw a boot on paper.

Hirsh-Pasek et al. (1984) found that children's ill-formed sentences were about twice as likely to be repeated by parents as well-formed ones. A person who denies the value of direct correction will point to the apparent approval conferred on the sentence by the adult. But the very fact that a sentence is repeated, let alone the intonation pattern used by the mother, singles the sentence out to the child as needing attention. Nor should the apparent infrequency of correction itself be a reason for dismissing it, a confusion of quantity with quality. The real argument once again is whether the type of knowledge postulated in UG is learnable through correction; in principle, this still seems as remote as ever. Chomsky originally stated: 'It is simply not true that children can learn language only through "meticulous care" on the part of adults who shape their verbal repertoire through careful differential reinforcement' (Chomsky, 1959, p. 42). Provided that the word 'language' is restricted within the UG scope, this seems still tenable. As a postscript, it should be noted that the results from a questionnaire given to parents in the Bristol project showed that, the more they believed they corrected, the slower the language development of their children (Wells, 1985, p. 351).

Turning to expansion, there has indeed been a widespread feeling that children benefit when adults expand their sentences. Bellugi and Brown (1964) described exchanges such as:

(30) Child: Baby highchair.
Mother: Baby is in the highchair.

as showing a process of 'imitation with expansion'; the mother expands the child's sentence and supplies anything missing while preserving the content words in their original order. However, Cazden (1972) reported a controlled experiment in which children whose sentences were expanded did *not* gain grammatically from the experience. Furthermore in a large-scale experiment, Nelson et al. (1973) found children who received expanded sentences progressed less well than those who received straightforward replies.

There are two main arguments why expansion is insufficient to acquire the aspects of grammar covered by the P&P Theory. One is the availability of expansions. Hirsh-Pasek et al.'s results are age-specific in that adults expanded the sentences of 2-year-olds but not those of 3- to 5-year-olds; at best only the early aspects of syntax could be acquired in this way. Wells (1985) found that parents expanded more when they knew the microphone was switched on: it may have been something they felt they ought to do rather than something they actually did.

The second argument against expansion would be the child's sheer difficulty in using expansions for acquisition. To appreciate that the adult is correcting a piece of syntax, the child has to disentangle the specific point from everything else included in the adult's expansion and to decide whether the adult is expanding a correct or an incorrect sentence. If expansion provides a type of negative evidence, it is not very efficient: a further example of I-language not being deducible from E-language without extra information.

5.2.2.3 *Social interaction*

The interaction between the child and the parents has often been seen in recent years as the mainspring of language acquisition. Correction, approval or imitation are different types of social exchange between the child and the parent; even if these do not carry sufficient weight separately, perhaps the child learns through a number of such routines. Jerome Bruner for instance attaches particular importance to 'formats'; a format is: 'a standardised initially microcosmic interaction pattern between an adult and an infant that contains demarcated roles that eventually become reversible' (Bruner, 1983, pp. 120–1), an example being the complex evolution of peek-a-boo games. Or take the following Bristol exchange:

(31) Father: Are you a mucky pup?
Child: No.
Father: Yes you are.
Child: No.
Father: Yes you are.

This seems a well-practised routine; the child may learn question formation in English by seeing how the question:

(32) Are you a mucky pup?

relates to the declarative sentence:

(33) Yes you are.

Many, if not most, researchers into child language since the late 1970s have connected the child's linguistic development on the one hand to the development of semantic meanings, on the other to social interaction. Bruner (1983, p. 34) talks of 'two theories of language acquisition; one of them, empiricist associationism, was impossible; the other, nativism, was miraculous. But the void between the impossible and the miraculous was soon to be filled in', in Bruner's view by showing how the child mastered 'the social world as well as the physical' (p. 39).

A more recent development in this vein bases social exchange through language on the speakers having a 'theory of mind'. In other words they have to attribute to other people the same ability to think and the same ability to show intentions and functions through language as they have themselves. To engage in interactive patterns with adults, children have to believe that adults have particular intentions in their minds behind their speech – an ability that seems to be missing in some autistic children. Take one of the elegant experiments described in Tomasello (1999). The child plays with the mother with three named toys; the mother goes out of the room and returns to find the child playing with a fourth unnamed toy; the mother exclaims 'Oh look! A modi!'; the child accordingly calls the toy a modi. The child has learnt what something is called without it being named. This can be explained only by attributing to the child a knowledge of the mother's intentions in speaking – she is talking about the new toy: the child has a theory of the mother's mind. Hence the approach is sometimes called 'theory theory' – the theory that people have theories of mind. Bruner's formats are then based upon shared attention and the understanding of another's intentions.

The arguments against social routines providing adequate evidence will be familiar by now: they provide positive evidence rather than negative evidence; there is a leap from familiar routines to the creative use of language. This is not to deny that such exchanges are vital for building up the use of language, pragmatic competence; 'it would not be at all surprising to find that normal language learning requires use of language in real-life situations, in some way' (Chomsky, 1965, p. 33). But UG Theory aims to explain grammatical rather than pragmatic competence; principles of UG are incapable of being learnt by social interaction. Whatever degree of importance one assigns to principles such as the Case Filter or binding, it is clear they are not learnable through such routines, however elaborate. A theory of mind may well be necessary for social interaction but this is not the aspect of language covered in UG, rightly or wrongly.

5.2.2.4 Dependence on other faculties

The other major alternative is that language acquisition depends upon general cognitive development. At stake is the autonomy of the language faculty. It is not denied that in actual use the production and comprehension of language depend upon other mental faculties and physical systems, although it is tricky to disentangle them. What *is* denied is that language acquisition depends upon other faculties. Piaget typically claims that the symbolic function of language depends upon the general semiotic function that develops out of the sensorimotor stage of cognitive development. The book *Language and Learning* (Piattelli-Palmarini, 1980)

provides a useful debate between Chomsky and Piaget on the issue of autonomy of language from other faculties. Chomsky points to the complexity of the knowledge that is learnt – phrase structure and the binding of *each other* – and denies that this could be the product of sensorimotor intelligence or of general learning theories. 'The common assumption to the contrary, that is that a general learning theory does exist, seems to me dubious, unargued, and without any empirical support or plausibility at the moment' (Chomsky, 1980b, p. 110). Typically he argues that, whatever else cognitive development can account for, it cannot explain the acquisition of language knowledge; as no one has proposed a precise way in which principles such as the Case Filter are acquired, they must be learnt in a manner specific to language. If one accepts Chomsky's premise that language consists of abstract principles such as the Case Filter, an attempt to show they are derived from other faculties must show not only their existence elsewhere but also how they are acquired, which he claims has not been done. The insufficiency of general cognitive development as a basis for language acquisition is demonstrated not so much by providing direct evidence as by challenging its advocates to show how specifically language knowledge is learnable by such means. 'No known "general learning mechanism" can acquire a natural language solely on the basis of positive or negative evidence, and the prospects for finding any such domain-independent device seem rather dim' (Hauser et al., 2002, p. 1577).

This discussion has sketched standard arguments against five alternatives to the UG position. The positions outlined are of course considerably more sophisticated than the brief versions given here. But overall the discussion has uncovered no way in which principles of UG are learnable from the environment. Figure 5.1 summarizes the argument, except for 'other faculties' which are hard to accommodate in the grid. Positive evidence alone is insufficient to acquire the principles of UG. The alternatives to innateness are insufficient because they rely on positive evidence, or they occur too rarely or too inconsistently, or they cannot explain creativity, or they cannot handle the type of knowledge that is acquired.

	Positive evidence	Other evidence	Occurrence requirement	Uniformity requirement
Imitation	+		–	–
Explanation		+	–	–
Correction		+	–	–
Social interaction	+		+	–

Figure 5.1 Some insufficient ways of acquiring Universal Grammar-based knowledge of the first language

UG Theory, however, is only concerned with core grammar, not with the many other aspects of language the child has to acquire. The arguments apply to the acquisition of the syntactic core: peripheral grammar, pragmatic competence, social competence, communication skills and so on may well be acquired by means such as the formats of Bruner. Indeed in some way these complement Chomsky's

approach by showing how the ability to use language may be acquired: 'The study of grammar will ultimately find its place in a richer investigation of how knowledge of language is acquired' (Chomsky, 1972b, p. 119).

5.3 The innateness hypothesis

So where does UG come from? Having eliminated all origins outside the mind, the only remaining possibility is that UG is already there. Important aspects of language are not acquired from experience: they are already present in the mind. 'The solution to Plato's problem must be based on ascribing the fixed principles of the language faculty to the human organism as part of its biological endowment' (Chomsky, 1988, p. 27). The distinctive quality of Chomsky's theory compared with other models is not innateness as such. Even a theory that children learn by associating pairs of words and objects attributes to them the innate ability to form such associations. 'Every "theory of learning" that is even worth considering incorporates an innateness hypothesis' (Chomsky, 1976, p. 13). The differences between language learning models lie in the nature and extent of the properties they attribute to the initial state. Chomskyan theory asserts that UG is the innate part of language knowledge: rather than a black box with mysterious contents, the mind contains UG principles and parameters.

The claim for innateness could be refuted in several ways. One is to deny the poverty-of-the-stimulus argument itself. But, so long as some aspect of language is known but not acquired from the world, the argument holds. A weaker attack is to deny that particular aspects of language are present in the final state of competence S_S, to reject, say, the Head Movement Constraint as part of the speaker's competence. However, the Head Movement Constraint is simply an explanation of why:

(34) *Is the teacher who here is good?

is ungrammatical; the only valid way of rejecting it would be to propose an alternative explanation; until that happens, it is the best explanation that is available. When a better proposal is made, it will be superseded. It seems doubtful whether any alternative principles that could account for this knowledge could be learnt either from positive evidence or from the likely negative evidence the child meets. The same is true of other principles. No one claims that the present principles represent the last word. But, to avoid the poverty-of-the-stimulus argument, the alternative principles would have to be learnable in one of the ways outlined above, which seems unlikely.

A further form of refutation would be to accept that a given aspect of language is present in S_S, the final state, but to demonstrate that it could in fact have been acquired from experience or from some other faculty in the mind. The Case Filter might be shown to arise naturally from some environmental factors, unlikely as this may seem. Or indeed it might come from other faculties of the mind. But dismissing a particular grammatical point from S_S does not defeat the argument itself, which could only be gainsaid by showing it applied to *no* aspect of language. Claims about innate ideas in UG Theory can always be found to be wrong; contrary evidence may show up and cause specific claims to be abandoned

or modified. 'An innatist hypothesis is a refutable hypothesis' (Chomsky, 1980b, p. 80). A theory based on evidence changes as more evidence comes to light, as do theories in other disciplines; UG Theory is not an unsubstantiable conjecture about the mind, but a hypothesis that is open to refutation and modification. Chomsky's argument that children are born equipped with certain aspects of language is justified by precise claims based on evidence about language knowledge; each piece of final competence that is not derived from experience is innate – the θ-Criterion, binding, or whatever. To defeat the argument involves explaining how each and all of these principles could have been acquired from experience or from other faculties. One ingenious way to circumvent the poverty-of-the-stimulus argument has indeed been to claim that language has become adapted over thousands of years by 'iterated learning' to the properties of the learner rather than the other way round (Zuidema, 2003).

5.4 The role of Universal Grammar in learning

Grammatical competence was presented earlier as knowledge of how the principles and parameters of UG are reflected in a particular language; knowledge of English is knowledge of how English utilizes UG. UG is the initial state of the language faculty, S_0, as we saw in chapter 2; it comes from within not from without. Acquiring English means discovering how it fleshes out the properties of UG which are already present. 'A study of English is a study of the realisation of the initial state S_0 under particular conditions' (Chomsky, 1986a, p. 37). The steady state is reached by the mind using evidence to discover how UG is instantiated in a particular language. When installing a new computer, though the modem is built in to the machine, it still needs setting to local circumstances – the appropriate phone number or connection and the server address. The final state S_S is one of the possibilities inherent in S_0, as are all the possible human languages; the contribution of experience is to decide which of these possibilities is actually realized. S_0 'projects' onto the final state S_S, as a frame in a film projects onto the screen. 'We can think then of the initial state as being in effect a function that maps experience onto the steady state' (Chomsky, 1980b, p. 109).

Children acquire movement in English by fitting what they hear to the pre-existing principle of Bounding in their minds. 'Language learning, then, is the process of determining the values of the parameters left unspecified by universal grammar' (Chomsky, 1988, p. 134). Children need to hear some examples of English sentences, or else they would have no reason for learning English rather than Japanese. The evidence encountered by the child need not be very extensive; a handful of English sentences could show how the head parameter applies, which words are anaphors, and that the language is non-pro-drop. Positive evidence 'triggers' acquisition rather than being needed in large quantities; some language is necessary to set the process off, to show how the principle applies or the parameter should be set. Chomsky (1988) pointed out that the single sentence 'John ate an apple' is enough to set the main word order parameters for English. But, unlike a language model, this experience does not form the primary source of information about language; a loud noise may trigger an avalanche but the cause

is the precarious state of the snow, not the accidental trigger. Experience sets off a complex reaction in the organism. Thus, although an I-language theory, UG has a place for experience in language learning – otherwise all children would end up speaking the same language. 'The environment determines the way the parameters of universal grammar are set, yielding different languages' (Chomsky, 1988, p. 134), or 'S_0 (=LAD) maps primary linguistic data (PLD) to L' (Chomsky, 2001a, p. 1).

Provision of appropriate input is completely necessary; indeed with certain constructions learning may take place faster if suitable triggering is provided; Cromer (1987) found that children who were given ten examples of the construction seen in *The wolf is easy to bite* every three months were, at the end of the year, on average ahead of those not given this slender exposure. It is not so much that the properties of the input have to be worked out by the child or that the grammar emerges in their minds as that properties of the input trigger appropriate parameter settings for lexical items to be stored in the mind.

UG cuts down the potentially infinite number of languages that could occur to the smaller number of possible human languages by imposing strong restrictions on their syntactic form. UG is a collection of restrictions on core grammar; the grammar of English consists of one combination allowed by these restrictions, the grammar of Chinese of another. Children narrow down the infinite possibilities of language to the one that they actually learn via UG. Given that a language could be anything at all, untold millions of children each year choose English, or German, or whatever language they are learning, out of the diverse possibilities; 'the system of UG is so designed that given appropriate evidence, only a single candidate language is made available' (Chomsky, 1986a, p. 83). Hence the reason why children learn language speedily, easily and uniformly – and apes and computers do not – is that UG narrows down the choices open to them and UG is not something that apes and computers possess.

EXERCISE 5.2 Categorize things that you can do by the different ways in which you acquired them.

	Imitation	Explanation	Correction	Social interaction	Innate
1 Typing					
2 Swimming					
3 Driving					
4 Breathing					
5 Reading					
6 Cooking					
7 Walking					
8 Dancing					

Do any of these have to be regarded as unlearnable and hence innate?

5.5 Complete from the beginning or developing with time?

One of the main issues about children's language can be seen in such children's sentences as:

(35) Mummy push baby.

Superficially this seems to be related to the equivalent adult sentence:

(36) Mummy is pushing the baby.

Such sentences as (35) are in fact typical of children at the two- and three-word stages. Sometimes they have been called 'telegraphic' since they resemble the old-fashioned telegram where payment was by the word, now perhaps more familiar as the text message constrained to 160 characters: children seem to be economizing on how much they can pack in. Sometimes their speech has been called 'semantic' since the sentences express an adequate semantic message with none of the grammatical trimmings, as seen in the small selection in figure 5.2.

Bike mummy.	There a sea.	Where write.
Crash up in.	Play race.	Me baby.
Haven't got.	Table up.	Stoppit bricks.
I loves you.	I wants a bit of milk.	Go down fat Pat's now.
Rachel like bottle.	Mummy make big noise.	

Figure 5.2 Some examples of English children's early sentences

The problem is how this stage relates to the adult grammar of linguistic competence towards which the child is heading. One possibility is that the child is indeed trying to express the adult sentence but something makes them miss out grammatical inflections such as *-ing* (*pushing*), articles such as *the*, and the auxiliary *is*, but leave in the content nouns (*Mummy, baby*) and the verb (*push*). This approach then sees children as having the entire adult grammar in their minds but being prevented in some way from expressing certain aspects. Over time the child is able to incorporate more and more of these elements of the whole grammar.

The alternative approach is to claim that the child has a unique grammar of their own that develops over time. At any point the child has a different grammar from the adult's, which gradually transforms over time into the final linguistic competence. Martin Braine (1963) was the first to show how the early two-word sentences of the child had a regularity of their own; he saw sentences such as *Bike Mummy* and *Look hole* as combining 'open' content words that could stand by themselves, such as *bike*, with 'pivot' words that only occurred alongside open words such as *look*. The child has a different grammar from the adult containing a single rule about how 'open' and 'pivot' words combine, not a skeleton of the adult grammar waiting to be fleshed out.

These two alternatives represent one of the perennial controversies in the field of L1 acquisition research: does the child start with the initial grammar more or

less in toto but not manifested in actual speech for one reason or another? Or does the child have a quite different grammar that gradually transforms into the final adult form? In terms of UG children might start with all the principles and parameters already present in their mind (but unable to express them in actual sentences) or the child might start with missing or different principles and parameters that change into the adult set. On the one hand is the **Full Competence Hypothesis**, on the other the **Gradual Development Hypothesis** (Borer and Rohrbacher, 2002). The controversy is echoed in SLA research in a slightly different form, as we will see in the next chapter.

It is clear that the controversial 'missing' elements from the children's sentences have certain things in common: they consist of functional categories such as complementizers and agreement inflections rather than lexical categories such as Nouns and Verbs. Elements that are often absent include:

- agreement markers such as third person *-s* from the IP:

  (37) Pig say oink.

- articles such as *the* and *a* from the DP:

  (38) Mummy push baby.

In terms of the English verb, this connects with the traditional grammatical distinction between finite verbs marked for tense, aspect, mood and voice etc., as in:

(39) The boy <u>worked</u> hard.

and non-finite verbs that are unmarked, as in:

(40) Work hard!

Or, in examples from the same boy, finite:

(41) I'm going down fat Pat's.

versus non-finite:

(42) Go down fat Pat's now.

Children who do not use verb inflections are going to produce sentences that coincide with adult non-finite forms. The issue is whether children already have a knowledge of the finite/non-finite distinction that is not evident in their surface use of verbal forms or whether they instead develop the distinction over time. The Full Competence Hypothesis insists that functional categories are present in the child's grammar from the beginning; the Gradual Development Hypothesis that functional categories are initially missing and develop over time.

One piece of evidence for the Full Competence Hypothesis is said to be the fact that when children *do* produce inflections they get them right, that is to say:

(43) The little thing goes round and round.

with the correct singular *-s*. Nevertheless it is easy to find counter-examples in the Bristol transcripts where the *-s* is used inappropriately, such as:

(44) I loves you.

In terms of syntax, the two alternative are whether the missing element is actually there somewhere in their knowledge of the grammar or is absent.

EXERCISE 5.3

What elements could you say were 'missing' from the following children's sentences? Do they support a view that the child is a deficient speaker of adult grammar or a view that the child is a speaker of an independent grammar of their own?

1 Gone down.
2 Tickling her.
3 All fall down.
4 Everythings car.
5 Me sit there.
6 Wash baby.
7 Dolly dress back on.
8 What doing?
9 Going to whish down in.
10 Mummy make big noise.

It is of course difficult to use a methodology founded on what is *not* present in children's language rather than what *is* present, called by Cook (1990) 'evidence of absence' – the same argument used by Sherlock Holmes when the dog *didn't* bark in the night. Arguing from what didn't occur is inherently problematic; there may be many reasons why the dog didn't bark as well as the fact that he knew the intruder, such as being asleep, preoccupied with a bone or drugged. Obviously there are many things children don't do at an early age for developmental reasons. Missing grammatical elements might be due to a processing deficiency – children's working memory is much smaller than adults' – or to perceptual problems – at least in English the functional elements alleged to be missing from children's language tend to be unstressed and lexical elements stressed. It is safer to use evidence of absence to support evidence of presence rather than to make everything hinge on what children *don't* say.

5.6 Issues in parameter setting

The classic case of 'missing' elements in children's sentences is the apparent lack of subjects in their sentences, as in:

(45) Haven't got.

Is this because they have no idea of subjects or because they can't express what they actually know? Chapter 3 described various ways of looking at the pro-drop parameter, which allows languages to have or not to have null subjects. How do children learn that Spanish is a pro-drop language, French is non-pro-drop, English is head-first, Japanese is head-last, and so on?

The classical GB Theory of the 1980s used the pro-drop parameter as a test-case for the acquisition of principles and parameters. Parameters in the child's mind can be thought of as built-in switches, each to be turned to suit the language that is heard. 'The transition from the initial state, S_0, to the steady state S_S, is a matter of setting the switches' (Chomsky, 1986a, p. 146). Acquiring the grammar of English means setting all UG parameters the English way; the setting of each switch is triggered by evidence.

To recap chapter 3, pro-drop languages allow sentences to have null subjects, as in Italian:

(46) È andato a scuola.

Non-pro-drop languages do not allow null subjects, so that the equivalent English sentence is ungrammatical:

(47) * Went to school.

Somehow Italian and English children acquire this divergent feature of their respective languages. This must depend on hearing sentences of either Italian or English. From this evidence alone, English children discover that English is a non-pro-drop language, Italian children that Italian is a pro-drop language. Somewhere in what children hear is the clue that tells them whether null subjects are allowed or not. Children must be learning either from positive evidence of actual sentences or from indirect negative evidence such as the lack of null-subject sentences in English. This process is possible only if their choice is circumscribed: if they know there are a few possibilities, say pro-drop or non-pro-drop, they only require evidence to tell them which one they have encountered. Again then parameters are a way of cutting down the choices the child could have to a manageable number.

Hearing a few sentences is sufficient to set the parameter one way or the other. The logic of indirect negative evidence, as Chomsky sees it, is: 'if certain structures or rules fail to be exemplified in relatively simple expressions, where they would be expected to be found, then a (possibly marked) option is selected excluding them in the grammar, so that a kind of "negative evidence" can be available even without correction, adverse reactions, etc.' (Chomsky, 1981a, p. 9). Indirect negative evidence does not circumvent the poverty-of-the-stimulus argument because it relies on the child's expectation of certain principles applying to the sentences he or she hears; in other words it presupposes innateness.

Three logical possibilities for parameters in the initial state S_0 were distinguished in GB Theory:

1 **The switch is in a neutral position.** The child is equally prepared for pro-drop or non-pro-drop settings. In this case the interim stages in the child's

development of grammar might have either setting; children learning Italian or English would not have a common sequence of acquisition but would set the pro-drop parameter appropriately from the moment that they first use it:

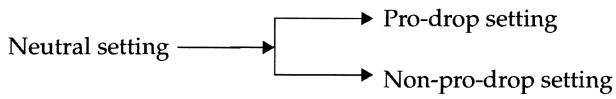


Figure 5.3 Neutral initial setting for pro-drop parameter

2 **The switch is set to non-pro-drop.** The child initially assumes that a subject is necessary whatever the language involved (non-pro-drop) and so needs evidence to set the parameter differently for pro-drop languages that allow null subjects. Children learning English would use one non-pro-drop setting from the beginning and would have no need to change it; children learning Italian would start with a non-pro-drop setting and would change it with time to pro-drop, triggered by evidence.

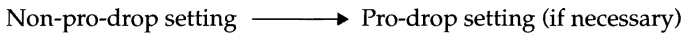


Figure 5.4 Non-pro-drop initial setting for pro-drop parameter

3 **The switch is set to pro-drop**, the reverse position. Those learning non-pro-drop languages are now the ones who require evidence that null subjects are *not* allowed. All children start from the pro-drop setting; those learning pro-drop Spanish need no extra evidence to reset it, those learning non-pro-drop English do.

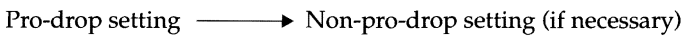


Figure 5.5 Pro-drop initial setting for pro-drop parameter

The most celebrated attempt to choose between these three possibilities for the pro-drop parameter in L1 acquisition was carried out by Nina Hyams (1986). By analysing published examples of young children's language, she found that English children at the earlier stages indeed produced null-subject sentences such as:

(48) Read bear book.

and:

(49) Want look a man.

These were not due to the children's limited capacity to handle information since at the same time they could produce equivalent sentences with subjects such as:

(50) Gia ride bike.

and:

(51) I want kiss it.

Nor was it a performance clipping of the initial *I*, as often occurs in adult speech, since the children also produced null subjects in other positions in the sentence. Hyams concluded that the third alternative is correct for pro-drop: children start from the setting that allows null-subject sentences whether they are learning English, German or Spanish. English and German children go on to learn that their languages are non-pro-drop, setting the switch away from its initial value; Spanish children can stick with the original setting.

But there is still no explanation how the child does this. Let us look at the evidence available to the child. It might seem that indirect negative evidence suffices to acquire non-pro-drop; noticing a lack of null-subject sentences, the child switches to non-pro-drop. However, this involves the child keeping track not just of those sentences that occur but also of those that do *not*: when, say, 500 sentences have been heard with nary a null-subject sentence among them, the switch reverses. Apart from the fact that this means ignoring the occasional subjectless sentence the child will hear for accidental or dialectal reasons, this would involve a striking feat of memory. Hyams (1986) presents a solution in which the child learns from positive evidence. This rests primarily on a further property of non-pro-drop languages within GB Theory described in chapter 3, namely the presence of 'expletive' subjects such as *there* and *it*. A non-pro-drop language such as English uses an impersonal indefinite *it* for instance in 'weather' sentences:

(52) It's raining.

Italian cannot use *it* in this way but must have a null-subject sentence:

(53) Piove.
(rains)

Similarly English has 'existential' sentences with *there* such as:

(54) There is a tide in the affairs of men which taken at the flood leads on to fortune.

which has a meaningless 'dummy' *there* in subject position. In other words it is a sign of a non-pro-drop language to have lexical expletives like *there* and *it* acting as subject; only languages with expletive subjects are non-pro-drop. Hyams (1986) suggests that the presence of such subjects constitutes one trigger that enables the child to set the parameter away from the initial pro-drop value. When English children hear:

(55) Once upon a time there were three bears.

or:

(56) It's time for bed.

they realize that English is non-pro-drop because of the dummy *there* and *it*, and this carries over to all the syntactic phenomena covered by the parameter. Hyams supports this by showing that English children acquire expletive subjects at about the same time as they acquire full lexical subjects. Thus the child is setting the switch from positive evidence alone, namely the use of *there* and *it* as subjects, rather than from the indirect negative evidence of never hearing any null-subject sentences.

Since 1986, Hyams has extended the analysis to other languages and substantially modified her position (Hyams, 1992). Her work provided an interesting insight into how the P&P Theory can be applied to actual data of children's language acquisition. The pro-drop parameter not only had theoretical linguistic consequences but motivated a clear analysis of one of the best-known aspects of children's early language, namely null-subject sentences. After Hyams's initial work, people could no longer say that UG Theory was divorced from the study of actual children's speech, spurring a whole generation of researchers to look at actual children's language from the viewpoint of principles and parameters.

As we saw in chapter 3, the analysis of null subjects has been modified by the notions of morphological uniformity and topic-drop, which would require children to set two parameters, not one. To settle whether topic-drop is a universally available option would mean showing children who have null subjects also have null objects. Data from young Chinese children showed indeed that they had both null subjects (46.5 per cent) and null objects (22.5 per cent) (Hyams and Wexler, 1993). On the other hand English children with null subjects (33.1 per cent) had only a few null objects (3.8 per cent) (Hyams and Wexler, 1993). This suggests that the two parameters are indeed independent: Chinese children have set the topic-drop parameter so that it covers both null subjects and null objects identified from discourse; English children have set the topic-drop parameter so that both null subjects and null objects are ruled out, but have still not set the pro-drop parameter to rule out null subjects from that source.

Using a limited number of principles like the Extended Projection Principle (EPP) and parameters like pro-drop, UG cuts down on the possible core grammars the child can choose from. Ideally, given the evidence that the child has available, UG should narrow the possibilities down to one: 'this language being a specific realisation of the principles of the initial state S_0 with certain options settled in one way or another by the presented evidence (e.g., the value for the head parameter)' (Chomsky, 1986a, pp. 83–4).

The evidence for fixing parameters need only be sufficient to trigger them and may be readily available. 'The parameters must have the property that they can be fixed by quite simple evidence, because this is what is available to the child; the value of the head parameter, for example, can be determined from such sentences as *John saw Bill* (versus *John Bill saw*)' (Chomsky, 1986a, p. 146). The effects of such parameters are sweeping; they are not confined to one rule or construction but apply anywhere in the grammar.

The discussion of language acquisition is thus no longer concerned with what happens in a single language; the interest lies in finding how the child's UG can cope equally well with different languages, or in fact *any* language. Most experimental or observational work with children has dealt with the acquisition of

particular rules in a language – how the child learns question formation, or passives, or relative clauses. From a UG perspective such work is at best partial. Rules such as question formation are not 'pure' discrete phenomena that can be studied in their own right but involve many principles, each of which has some contribution to make. Research based on rules or on specific constructions is only a first approximation. Instead UG-based research needs to examine how a principle or a parameter is employed across the board in the child's grammar before it can make a final statement about acquisition. The examples given here have indeed run some risk of being construction specific; for example the discussion of null-subject sentences may seem to concern a single sentence type rather than an underlying parameter that manifests in other aspects of the sentence as well as the subject position.

EXERCISE 5.4 Here are some sentences of Arabic with English literal translations. From these (i) is Arabic a pro-drop or non-pro-drop language? (ii) would this evidence be sufficient for the child to acquire the right pro-drop setting for Arabic?

1. Howah fi el-hograh el-okhrah.
   He (is) in the other room.

2. Honakah tofaha fi el-matbakh.
   There (is) an apple in the kitchen.

3. Heiah frensiah.
   She (is) French.

4. Ennaha tomter.
   It (is) raining.

5. RaaH al beet.
   (He) went-3rd.pers.sing home.

6. Men el-sa'b an-toghany.
   It (is) hard to sing.

7. Honakah onas fi hadikatonah.
   There (are) people in our garden.

8. Badde tiffeeHa.
   (I) want-1st.pers.sing apple.

9. Sayathhaboona ela el-senima.
   (They) will go to the cinema.

5.7 Markedness and language development

UG is concerned with core grammar rather than with the periphery. The logical argument of acquisition therefore deals primarily with the core – the elements that are directly linked to UG. Peripheral elements can be learned in ways that that are unconnected to UG. Politeness formulas such as *please* and *thank you* may well be learnt through active correction by the parents; historical accidents such as the irregular past tense *dove* in English are not necessarily learnt through UG or entirely from positive evidence. The same idealizations are involved in acquisition as in the description of competence; it is grammatical *knowledge* that is being discussed, not language performance or language use; within language knowledge, the crucial areas are those that are universal.

Inside the core grammar, one possibility is that certain parameter settings are more marked than others; languages that have overt syntactic movement, for instance, may be closer to the unmarked form of UG than those that do not, making English unmarked, Japanese marked. With the opposite assumption, English is marked, and Japanese is not. Going back to pro-drop, the conclusion that all children start with a pro-drop setting means that non-pro-drop is more marked. Children start off with the unmarked setting for the parameters; they have to reset those which are more marked in the language they are learning.

Markedness also relates to the problem of evidence available to the child. One interpretation is that the unmarked settings of parameters are those that the child can learn from the least amount of positive evidence. Italian learners need no evidence to set pro-drop; they have the right setting from the start. Children learning English need evidence to turn the switch away from its initial unmarked setting, if we accept Hyams's account. Children need evidence to move from unmarked to marked settings. Hence marked elements of UG need more evidence, or different types of evidence, than unmarked elements. Elements of peripheral grammar may need totally different types of evidence; for example the *was/were* distinction in English might need specific negative correction from parents. Perhaps it should be pointed out that the use of the term 'markedness' in UG is very different from its use in other theories.

Complementary to this approach to markedness is the Subset Principle, introduced within P&P Theory by Berwick (1985) and stated by Chomsky as: 'if a parameter has **two** values + and –, and the value – generates a proper subset of the grammatical sentences generated with the choice of value +, then – is the "unmarked value" selected in the absence of evidence' (Chomsky, 1986a, p. 146). Roughly speaking, children choose the setting for a parameter that fits the evidence with the fewest possible assumptions. The children's choice is conservative in that it stays as close as possible to the data they hear. They prefer a language that is a 'subset' of a larger language rather than leaping immediately to the 'larger' version. This is slightly difficult to accommodate within the discussion here since it concerns the 'languages' the children learn rather than 'grammars'. Wexler and Manzini (1987) show how the Subset Principle deals with the learning of the binding principles in terms of wider and wider inclusive definitions of local domain

but argue it does not apply to the type of evidence for pro-drop presented by Hyams (1986).

There are general problems in interpreting data from the actual language development of children in relation to markedness. The study of children's speech is potentially misleading for the logical problem of acquisition. As with adults, grammatical competence is imperfectly reflected in speech performance; children too can run out of breath, or make mistakes or change their minds. The psychological processes used in speech comprehension and production are indirectly and partially linked to their grammatical competence. To study the linguistic competence of adults, an alternative source of evidence is available in the form of their judgements about sentences, a device used frequently in this book. Such judgements are hardly feasible with small children. Nevertheless some researchers have argued in its favour: McDaniel and Cairns (1990) for instance suggest that judgements can be elicited from small children if the task is appropriately designed: Hannan (2004) adapted the technique to bilingual children as young as 4 years old. However large the sample of children's utterances, it is still an inaccurate source of information about their competence.

In addition the processes involved in language performance are themselves developing at the same time as competence. Filtering out the effects of performance processes from actual samples of speech is doubly difficult with children since it cannot be assumed that it is only language that develops. The child starts by saying one word at a time:

(57) Mine. Bath. Yes. Car. Carses. Hiding.

and goes on to a two-word stage:

(58) That baba. All gone. A lion. Little girl. See Mary.

Both stages may be the by-product of short-term memory restrictions that limit the number of items in the child's utterance rather than of anything directly to do with language acquisition; 'it might be that he had fully internalised the requisite mental structure, but for some reason lacked the capacity to use it' (Chomsky, 1980a, p. 53). The apparent progress from one-word utterances to two-word utterances may have little to do with language acquisition as such, more to do with the growth of 'channel capacity'. The expansion of children's general cognitive capacity allows them to produce longer and more complex sentences, but this is caused by relaxing constraints on performance rather than by their increasing competence. In Chomsky's words: 'much of the investigation of early language development is concerned with matters that may not properly belong to the language faculty . . . , but to other faculties of mind that interact in an intimate fashion with the language faculty in language use' (Chomsky, 1981b, p. 36), i.e. in terms of chapter 2, the broad faculty of language (FLB) rather than the narrow faculty of language (FLN). Acquisition considered as a logical problem is an abstraction from such features of development. The history of speech developing in the child reflects factors that are nothing to do with acquisition.

We can make a distinction between language **acquisition** – the logical problem of how the mind acquires S_S independent of intervening stages – and language **development** – the history of the intervening stages (Cook, 1985); the distinction reflects Chomsky's thinking, even if he does not use these terms. Acquisition is an idealized 'instantaneous' model in which time and experience play minimal roles; the crucial factor is the relationship between the initial S_0 and the steady-state S_S. Development reflects the complex interaction of language with the other faculties of mind that are maturing at the same time. Research into both acquisition and development is concerned with evidence, the former with evidence of what speakers know, the latter with evidence of what children say. Acquisition theory does not necessarily need support from actual studies of children's language; 'behavior is only one kind of evidence, sometimes not the best, and surely no criterion for knowledge' (Chomsky, 1980a, p. 54). While some studies such as Hyams (1986) have attempted the complex task of linking development with acquisition, there is no compelling reason why the theory should accept this as more than supplementary evidence.

It is possible to assign markedness as a post hoc consequence of developmental studies. If the pro-drop setting is used first, then, everything else being equal, this is one reason for preferring pro-drop as the unmarked setting. But everything else rarely *is* equal because of the complexity of actual development. 'We would expect the order of appearance of structures in language acquisition to reflect the structure of markedness in some respects, but there are many complicating factors: e.g., processes of maturation may be such as to permit certain unmarked structures to be manifested only relatively late in language acquisition, frequency effects may intervene, etc.' (Chomsky, 1981a, p. 9). Assigning markedness on the basis of developmental stages is circular if there is no other reason for a setting to be unmarked than its earlier occurrence; without a syntactic rationale, markedness amounts to saying 'whatever is learnt first is learnt first'.

It might also be that the language faculty itself matures. Babies have lungs and hearts that function from the time that they are born; however, the first teeth appear at around 7 months, and are replaced by others at around 6 years; wisdom teeth appear much later. This does not mean that teeth are not biologically determined; they appear at particular stages of maturation even if absent at the beginning; 'genetically determined factors in development evidently are not to be identified with those operative at birth' (Chomsky, 1986a, p. 54). UG might be like the heart – complete and functional at birth – or like teeth – coming into operation bit by bit; these alternatives are called by Chomsky the 'growth' theory and the 'no growth' theory (Chomsky, 1987), by others the Gradual Development Hypothesis (Déprez, 1994) and the Full Competence Hypothesis (Poeppel and Wexler, 1993). So far as acquisition is concerned, UG is neutral between the two possibilities; the relationship between the two states, S_0 and S_S, is unaffected by whether UG is initially present or not. It is, however, vital to the development argument. On the one hand in the no-growth version all the apparatus of the computational system complete with principles and parameters is present from the very beginning, waiting to be fleshed out through the language input that it encounters; any differences between the child's grammar and the adult's competence are a matter of not having acquired particular information from the input to set parameters. On the other

hand in the growth version the computational system is only partly present initially and develops with time; differences from adults may be lack of the actual principles or parameters as well as lack of parameter setting.

Let us borrow the term 'wild grammar' from Helen Goodluck (1986) to refer to a grammar that does not conform to UG, by, say, breaking the principles or having illegal parameter settings. If UG is present and functioning from the start, children will never entertain wild grammars; children will not stray outside the bounds of UG at any stage of development; their learning will be error-free so far as UG is concerned. A UG principle such as Relativized Minimality will be used at all stages of development. This is not to say it figures prominently in their speech from the start; it may be precluded for performance reasons because the sentences they can say are not long enough or not complex enough; a principle of binding for example can hardly be used in one-word or two-word sentences. But, if there are no wild grammars, none of the children's grammars should violate Relativized Minimality or binding at any stage; all of their interim grammars should be possible human languages. Research into language development to investigate this is necessarily complex and tentative. Anecdotal observations suggest that children rarely produce sentences that breach UG. But this is hardly surprising in view of the complexity of the sentences that are needed to show many UG principles, combined with the constraints on the child's performance; there are no opportunities for visibly breaking bounding with one-word sentences, for example.

It is perfectly plausible that UG matures. Lila Gleitman has often argued for a biologically determined maturational transition from a semantic phase of language to a syntactic phase; at the first stage the child produces sentences to convey meaning without regard to their syntax; at the next stage the child abruptly switches to syntactic organization. Gleitman's evidence is based on a variety of forms of acquisition including acquisition by Down syndrome children, blind children, deaf children and premature children (Gleitman, 1982). Chomsky's view that early stages of development are nothing to do with acquisition supports the possibility of a pre-syntactic phase; UG may simply not be available to the very young child, who gets by on semantics alone. Chomsky separates this line of thinking firmly from the logical problem of language acquisition; 'there is good reason to believe that the language faculty undergoes maturation – in fact, the order and timing of this maturation appear to be rather uniform despite considerable variation in experience and other cognitive faculties – but this does not bear on the correctness of the empirical assumption embodied in the idealisation to instantaneous learning' (Chomsky, 1986a, p. 54).

Let us review briefly the main argument for the innateness of UG. The first step is to recognize the complex and abstract grammatical competence possessed by the native speaker. The distinctive nature of this knowledge rules out its acquisition through imitation, correction and approval, social routines, or other mental faculties; grammatical explanation is ruled out for reasons of probability and lack of universality. Unless the principles and parameters of grammatical competence can be shown to be learnable by one or other of these means, they must be innate. This central argument is bolstered by other arguments about the common possession of language by the whole human species, the uniformity of

acquisition despite the variety of situations, the lack of key mistakes in children's speech, and the inability of other species to acquire language. But the crucial step is the first one: once it is conceded that language knowledge is defined in terms of a grammatical competence of this kind, everything else follows.

How does it relate to other approaches to language acquisition? The UG position that has been presented is, broadly speaking, popular among linguists rather than among those primarily interested in studying L1 acquisition in children. One reason for this is that, in a sense, the study of actual children is unnecessary in UG Theory since all that the linguist needs to do is to work backwards from S_S, the final state. The basic work is the linguistic description and the use of learnability arguments to motivate the description. E-language research with large numbers of children is secondary even if it can provide some supporting evidence, if properly evaluated. I-language research deals with the individual mind, not the behaviour of groups.

Secondly, experimental or observational research on UG issues is hard to carry out. The necessity of separating competence from performance and acquisition from development means on the one hand that actual observations of children are flawed, on the other that appropriate methodologies for experiments are extremely hard to devise.

The impression one is left with is that many child language researchers are not prepared to take the first step of accepting that the essential part of language is grammatical competence. They may concede that it is a curious logical problem, but claim that what really interests them is how children form social relationships, how their thinking influences their language, how children with language problems can be helped, all side-issues to UG Theory. To those with primarily sociological or educational aims, or indeed E-language aims in general, UG Theory has relatively little to offer; it is concerned with mental man rather than with social man, and with what human beings have in common rather than their differences. UG indeed provides a core test case of the essential quality of human minds; the type of knowledge revealed as part of the structure of the brain is fascinating and profound. Supporting it with research into children's language is a valuable exercise. UG Theory has a unique central place in L1 acquisition studies. But it is only part of the broad picture. UG Theory is concerned with the acorn rather than with the tree in all its complexity; vital as the acorn may be as the source of growth and development, for many purposes the leaves, the wood or the blossom are more important. The danger is that UG may be seen as a threat to other ideas of language development, rather than as a complementary theory that accounts for a specific area of vital concern to those interested in the uniqueness of the human mind.

This discussion has drawn on Chomsky's ideas of language over five decades. In terms of syntactic acquisition perhaps the major breakthrough was the principles and parameters model which managed to unify the description of syntax with the model of acquisition. The Minimalist Program (MP) has so far had much less impact on ideas of acquisition, even after some fifteen years of development. The central ideas seem murky in an acquisition context and have so far yielded little new insight into acquisition. Most L1 researchers seem quite happy to continue using a version of the P&P Model that involves little of the MP.

6 Second Language Acquisition and Universal Grammar

A large proportion of the human race, some would say the majority (Cook, 2002), speak more than one language. Somehow two languages, two grammars, can coexist within the confines of one mind. Is the existence of so many minds with two or more languages at all relevant to the Universal Grammar (UG) Theory? This chapter extends the discussion of UG to second language acquisition (SLA). Like chapter 5, it does not attempt to give a complete survey of the extensive research in this area but aims to review the issues and ideas. For a more detailed and comprehensive treatment readers are referred to White (2003).

6.1 The purity of the monolingual argument

According to the separation of competence from performance, discussed in chapter 1, 'Linguistic theory is concerned with an ideal speaker-listener in a completely homogeneous speech community' (Chomsky, 1965, p. 4). A community with more than one language, or indeed more than one dialect, would not be homogeneous: the language of a mixed community 'would not be "pure" in the relevant sense, because it would not represent a single set of choices among the options permitted by UG but rather would include "contradictory" choices for certain of these options' (Chomsky, 1986a, p. 17). The idealization of competence reduces it to the knowledge of a monolingual native speaker. Describing a mind with two grammars is too complicated; it is vital to simplify the discussion to a mind with a single grammar. At the level of descriptive adequacy, the goal is therefore to describe what an idealized monolingual knows: the description of UG is based on the knowledge of the native speaker with a single grammar.

Mostly this emphasis on monolingualism has simply been taken for granted by those working within the UG Theory, along with the other areas excluded from competence, and is seldom discussed or justified. The only true knowledge of the language is taken to be that of the adult monolingual native speaker. Chomsky himself has rarely mentioned bilinguals. In a famous interview with François Grosjean, he said:

> Why do chemists study H_2O and not the stuff that you get out of the Charles River? . . . You assume that anything as complicated as what is in the Charles River will only be understandable, if at all, on the basis of discovery of the fundamental principles that determine the nature of all matter, and those you have to learn about by studying pure cases.

(For the benefit of readers without a knowledge of US geography, the Charles River divides Boston from Cambridge.) This can be called the 'purity' argument for using monolinguals: an idealized mind with one language possesses a purer state of language knowledge than a mind with two or more.

Chomsky does not of course deny that there are large numbers of bilinguals in the world: 'even in the United States, the idea that people speak one language is certainly not true . . . everyone grows up in a multilingual environment' (Chomsky, 2000a, p. 59). But this is irrelevant to the idealized competence of an individual. A second language (L2) is effectively an extra tacked on to the first language (L1), like an extension to the back of the house. Doubtless it has an interest of its own in due course. But such an impure state cannot form the core subject-matter of linguistics.

6.2 Universal bilingualism

Alongside the purity argument, there is nevertheless a recognition that many, or indeed all, minds contain more than one grammar: 'whatever the language faculty is it can assume many different states in parallel' (Chomsky, 2000a, p. 59). Chomsky has often made comments to the effect that people effectively have more than one grammar in their minds: 'every person is multiply multilingual in a more technical sense' (Chomsky, 2000a, p. 44). For example, we may switch between speaking different dialects with different parameter settings, say standard English *She's good* versus dialects such as US Black English that permit sentences without copula *be* as in *She good*. Or we may use different parameter settings for different registers: Haegeman and Ihsane (2002) have shown that diary writing in English in writers such as Virginia Woolf (or the fictional Bridget Jones) is pro-drop in that the first person is often a null subject – *played gramophone . . . so to tower* – not a characteristic of Woolf's public prose style. Hence she is switching between two settings for the two styles and so has two grammars simultaneously.

A person who switches in this manner then has elements of two grammars in one mind. Thomas Roeper argues that: 'a narrow kind of bilingualism exists in every language. It is present whenever two properties exist in a language that are not statable within a single grammar' (Roeper, 1999, p. 169), for example when adults with 'optional' rules are switching between two grammars as in the case of Woolf. Chomsky too often talks of bilingualism proper as simply the extreme end of a continuum of grammatical variation inherent within all speakers. 'To say that people speak different languages is a bit like saying they live in different palaces or look different, notions that are perfectly useful for ordinary life, but are highly interest-relative. We say that a person speaks several languages, rather than several varieties of one, if the differences matters for some purpose or interest' (Chomsky, 2000a, pp. 43–4).

The simultaneous existence of two grammars in the same mind is also necessitated by language development (Roeper, 1999). It might be that children switch in toto from one parameter setting to another so that, say, one day they have null subjects in their speech, the next day they do not. Studies of development, however, show that such an abrupt transition seldom occurs: English children gradually decrease the number of null subjects in their speech over a period of time rather than going from having no subjects at all to having them in every sentence. Hence, during this transitional period, they must in effect have two grammars simultaneously, or at least two parameter settings. Any transition from one stage to another involves bilingualism in the sense of knowing two grammars or having two sets of parameter settings for an appreciable amount of time. The abstraction of competence to a single grammar is a fiction for most L1 native speakers who can use different dialects or genres and for most L2 learners; the typical human mind must entertain more than a single grammar. The issue is really whether it is proper to set this universal bilingualism to one side in linguists' descriptions of competence or whether it should in effect form the basis of the description from the beginning.

EXERCISE 6.1

Here are some sentences by (i) native speakers of standard English, (ii) non-native speakers of English, (iii) dialect speakers of English (represented by the spelling). Judging from the syntax alone try (a) to classify them into the three groups and (b) to see which of them are hard to understand.

1 Wot sort of party was this you was boaf at, anyway?
2 I hain't hearn 'bout none un um, skasely, but ole King Sollermun, onless you counts dem kings dat's in a pack er k'yards.
3 They didn't said exactly what is it for, but for what they told me it's to give all of us a great notice.
4 It's the guardian angel of twocers. It's like wor logo, y'kna.
5 I have wrote to Dennis but I haven't received any answer yet.
6 I hate smoothies. Too fattening. My vice is pizza. Next time bring me a pizza.
7 Stillman's face. Or Stillman's face as it was twenty years ago. Impossible to know whether the face tomorrow will resemble it.
8 I think that one of the only ways to prevent the worse fear which can striken you, that is, the fear of the unknown, is to communicate with others.

Does this mean that the norm for a language is the monolingual native speaker rather than speakers of dialects or L2s?

6.3 The multi-competence view

The multi-competence theory, within which one of us works (Cook, 2002), sees these issues differently. If most people, or indeed *all* people, have multiple grammars in their minds, the idealization to the monolingual native speaker is misleading, as inaccurate as studying how human beings breathe by looking at those with a single lung. If the architecture of the human mind involves two

languages, we are falsifying it by studying only monolingual minds. To turn Chomsky's metaphor back on him, water is a molecule, H_2O, not an atom; if we break it into its constituent hydrogen and oxygen, we are no longer studying water. Purifying the mind into a single language means destroying the actual substance we are studying – the knowledge of language in the human mind.

The arguments about language acquisition discussed earlier were couched in terms of the potential inherent in all human children irrespective of the environment they encounter. Following the same line of reasoning, potentially all children can become bilingual; the ability to know more than one language is available to us all, even if it may decline after childhood. A person is a monolingual because of the accidental fact that they only encountered one language, and were unable to realize their bilingual or indeed multilingual potential. If you don't hear an L2, you won't speak one; if you do, you will. Since every human being has the potential to do this, UG Theory has to take multilingualism as the norm for the human mind. Multiple grammars in the mind are not the exception but the norm, prevented only by accidental environmental features.

According to the multi-competence theory, then, the linguistic competence in the human mind potentially includes more than one language. UG Theory has to account for this universal ability of the mind to have two, possibly conflicting, grammars at the same time; universals cannot be established by studying the minds of people who know one language, only minds of people who have fulfilled the multilingual potential of the human language faculty. Rather than making L2 user grammars conform to the Procrustean bed of the monolingual grammars, we need to establish principles from L2 grammars and see monolingual grammars as a restrictive set.

This position then treats the multilinguals of the world as the norm, not the monolinguals. The inhabitants of the Cameroon for example use 279 indigenous living languages (Gordon, 2005). A priest from Tanzania spoke Kihaya as a child, learnt Kiswahili in elementary school and English in secondary school, needed Latin for his religious training (but also learnt French out of curiosity at the same time), was posted to Uganda and Kenya, where he needed Rukiga and Kikamba, and went to Illinois where he uses Spanish to communicate with his parishioners (Cook, 2001). Londoners speak over 300 languages and 32 per cent of their children have languages other than English at home (Baker and Eversley, 2000). While these seem extreme cases to those brought up as monolinguals in households with one language, they show that the potential for multi-language acquisition is dormant within everyone, even in countries such as England that are supposedly monolingual – exactly the kind of insight that the UG Theory was set up to explain.

6.4 The poverty-of-the-stimulus argument and second language acquisition

How does the poverty-of-the-stimulus argument apply to L2 acquisition? If L2 learners possess knowledge of language they could not have acquired from the evidence they have encountered, its source must be within their own minds: the

same logic applies as to L1 acquisition. Innateness can be established in the same fashion in L2 learning as in L1 acquisition.

The poverty-of-the-stimulus argument, however, works slightly differently with L2s from the steps presented in chapter 2 (p. 55). We will distinguish the steps in the L2 argument from the L1 argument with primes – A′, B′, etc. Step A in L1 acquisition (p. 55) means demonstrating the existence of a property of syntax in the mind of a native speaker, taken as normal by definition. But L2 learners come in all varieties and levels of knowledge of the L2: some are just beginners and never likely to progress any further; others are interpreters for the UN with the future of nations hanging on their translations. There is no typical L2 learner, only diverse individuals. Hence Step A′ in L2 acquisition already has to be qualified.

Step B′ of the argument means showing that this aspect of syntax is not acquirable from input available to the L2 learner. The occurrence and uniformity requirements described in chapter 5 function slightly differently in L2 learning theory, since the final L2 knowledge is itself variable. Explaining how a learner acquired something means showing that the postulated situational effect actually occurs for that learner, but does not necessitate showing it occurs for all learners – what are termed elsewhere the 'narrow' and 'broad' forms of the poverty-of-the-stimulus argument (Cook, 1991).

6.4.1 *The evidence available to the second language learner*

To demonstrate that Step B′ applies to SLA, we can review the same strategies that were applied to L1 acquisition in the last chapter.

6.4.1.1 *Imitation*

As in L1 acquisition, sheer imitation only provides positive evidence of what is heard: repeating sentences does not in itself allow the learner to know what can*not* be said. Indeed a similar argument was responsible for the decline of language teaching methods that rely on imitation, such as audiolingualism (Lado, 1964). Repeating aloud:

(1) Oscar fancies himself.

ten times does not confer knowledge that *Oscar* and *himself* refer to the same person.

6.4.1.2 *Direct teaching: explanation*

Grammatical explanation does not figure prominently in the experience of L1 children. L2 learning may, however, be different, at least for those learners who encounter the language in the classroom. Some language teachers constantly explain rules to students; explanation is a corner-stone of the grammar-translation method still popular in many European teaching settings, in particularly at university level (Coleman, 1996). In the 1990s there was a limited return to grammar explanation in language teaching through the 'focus on form' technique that became the distinctive element of task-based teaching (Willis, 1996). Carroll and Swain (1993)

showed that 'metalinguistic feedback' indeed helped L2 learners to acquire the 'dative alternation' variation between *Give the dog a bone* and *Give a bone to the dog*. Even if some learners never encounter grammatical explanation, undoubtedly many adult L2 learners automatically reach for a grammar book.

Yet the explanations of syntax that L2 learners receive necessarily concern only those points that their teachers are aware of; Relativized Minimality or the null subject are unlikely to be part of most teachers' conscious grammatical knowledge. 'It must be recognised that one does not learn the grammatical structure of a second language through "explanation and instruction" beyond the most elementary rudiments, for the simple reason that no one has enough explicit knowledge about this structure to provide explanation and instruction' (Chomsky, 1969, reprinted in Chomsky, 1972a, pp. 174–5).

Even if grammatical explanation might work for some aspects of L2 learning, it cannot account for how people know what they are *not* taught, for example the Locality Principle. Take the explanation of 'reflexives' in a typical pedagogical grammar book: 'pronouns ending in self and selves are used when we want to say that the subject of the sentence does (did/has done, etc.) something to itself' (Bald et al., 1986, p. 111). This hardly goes very far towards explaining the binding principles the L2 learner needs. This does not exclude the possibility that L2 teaching could hypothetically be based on grammatical explanation of principles and parameters syntax, as suggested in Cook (1993) and carried out with some success in White (1991).

6.4.1.3 Direct teaching: correction and approval

Children rarely receive correction or approval of syntactic forms in L1 acquisition; the occurrence requirement therefore rules this out as a way of acquiring the L1. But such feedback is provided in many L2 learning situations, most conspicuously in the classroom, but also in 'natural' situations; teachers are only too aware of the insatiable demand from L2 students for correction. If correction is to be successful, the L2 learner must, furthermore, produce sentences that deviate in the appropriate way. In order to learn the Head Movement Constraint, the learners must produce sentences such as:

(2) Is Newcastle is the city that on the Thames?

And the correctors must point out:

(3) No you must say 'Is Newcastle the city that is on the Thames?'

refraining carefully from pointing out Newcastle is on the River Tyne. Though the language of L2 learners exhibits a variety of peculiarities, such mistakes are not numbered among them. It is not obvious that L2 learners actually produce the necessary mistakes in terms of principles and parameters that would enable their teachers to correct them.

The corrector has to be able to identify the problem in order to correct it. Many of the possible deviations from UG are unlikely to be spotted by the ordinary native speaker. Correction cannot of course be ruled out as a source of evidence

in the classroom: traditionally minded teachers use it frequently; conventionally minded students often request it. But correction of the type of mistake needed to acquire UG principles seems unlikely. Correction is no more likely to lead to L2 knowledge in non-classroom L2 settings than in L1 acquisition. As L2 learners manage to learn UG principles without such correction, it cannot be the most important element. While correction potentially meets the occurrence requirement for *some* areas of language for *some* learners, it is unlikely to prove an effective way of acquiring the central areas of UG.

6.4.1.4 Social interaction

To look at social interaction in L2 acquisition means separating those exchanges that are 'natural' from those that are 'non-natural'. Though L2 learners may engage in the same 'natural' routines of social interaction as L1 children, they may in addition have controlled 'non-natural' exchanges, for instance those found in teaching techniques such as structure drills in which a single grammatical pattern is practised over and over. Natural social exchanges seem to provide a clear route to pragmatic competence in the L2 but they are not able to facilitate the acquisition of UG principles in the L2 any more than they are in the first. Indeed, while all L1 learners encounter natural communicative interaction, many classroom L2 learners do not.

Non-natural exchanges have been used by teachers in ways that range from grammatical correction to asking the students to talk about the differences between two pictures to the classic three-fold exchange described as IRF – Teacher's Initiation (I) / Pupils' Response (R) / Teacher's Feedback (F) (Sinclair and Coulthard, 1975). A popular topic in recent classroom-based L2 research has been the use of 'recasts' in the classroom – occasions when the learners' mistakes are paraphrased by the teacher, surveyed in Han (2002), similar to the 'imitation with expansion' suggested by Bellugi and Brown (1964) for L1 acquisition. These exchanges could aim at teaching principles such as binding through a carefully constructed Socratic dialogue in which the student is led to see the binding possibilities in the sentence. Whatever other aspects of language these exchanges may promote, to our knowledge this approach has not been attempted to teach UG principles, and would indeed necessitate the provision of negative evidence or grammatical explanation if it were to succeed. Vital as social interaction may be to the overall needs of foreign language students, it is an unlikely vehicle for the acquisition of core UG syntax.

6.4.1.5 Dependence on other faculties

The use of other mental faculties was ruled out in L1 acquisition, primarily because of the uniqueness of the language principles. The same argument applies to L2 learning; there is no compelling reason why other mental faculties should be involved. However, the L2 argument is more complex as the L2 learner is usually at a later stage of cognitive development than the L1 child. Consequently the relationship between language and cognition differs from that in the native child. This implies that the knowledge of the L2 differs from the knowledge of the L1, and so is not expressible in the Principles and Parameters (P&P) format (since these cannot be learnt).

	Positive evidence	Other evidence	Occurrence requirement	Uniformity requirement
Imitation	+		±	–
Explanation		+	±	–
Correction		+	±	–
Social interaction	+		±	–

Figure 6.1 Some insufficient ways of acquiring Universal Grammar-based knowledge of the second language

We have then shown that Steps A and B of the poverty-of-the-stimulus argument apply to L2 acquisition, with some qualifications in that there may not be a final steady state of L2 knowledge and that situations of L2 acquisition vary in ways that those of L1 acquisition do not, thus making the uniformity argument hard to apply. The remaining two steps of the poverty-of-the-stimulus argument then follow. Having demonstrated that learners know something (Step A′) that they could not have acquired from outside (Step B′), we can argue that, at least in a large number of cases, this aspect of syntax is not acquired from outside or transferred from the L1 (Step C′) and thus must be built in to the learner's mind (Step D′). These steps with their modifications are given in figure 6.2; they parallel the L1 steps given on page 57; the prime after each letter reminds us that none of the stages A′–D′ is precisely the same as the A–D in L1 acquisition.

Step A′: An L2 user of a particular language knows a particular aspect of L2 syntax.
Step B′: This aspect of syntax could not have been acquired from the language input typically available to L2 learners.
Step C′: We conclude that this aspect of L2 syntax is not learnt from outside.
Step D′: We deduce that this aspect of L2 syntax is built in to the mind.

Figure 6.2 The poverty-of-the-stimulus argument applied to second language acquisition

All in all, the poverty-of-the-stimulus argument still applies to L2 acquisition, provided various provisos are borne in mind about the alternative starting and finishing points and the routes between, to be discussed below. If some L2 learners who have used natural means know UG principles such as the Case Filter, the source must be in their own minds. But of course this may reflect the knowledge they already have of an L1, rather than UG itself.

6.5 Models and metaphors

We can now attempt to see how the states and Language Acquisition Device (LAD) metaphors presented for L1 acquisition in chapter 2 can be applied to the acquisition of other languages than the first.

6.5.1 The Language Acquisition Device metaphor

In principle the LAD/UG metaphor used for L1 acquisition could be extended to take in other languages. A second set of primary linguistic data goes into the black box; a second grammar comes out containing a second version of the principles, a second batch of settings for the parameters, and a second lexicon, yielding the model seen in figure 6.3.

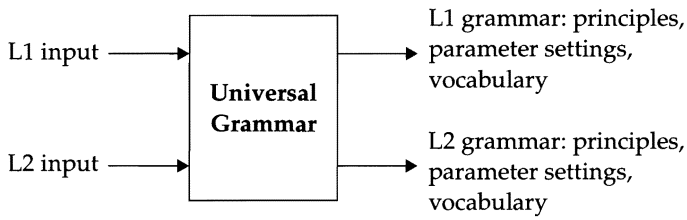


Figure 6.3 The Universal Grammar model of first language acquisition extended to second language acquisition

In L1 acquisition, virtually all children acquire full L1 competence. Though many people start to learn an L2, few, if any, manage to gain a knowledge of the L2 equivalent to their knowledge of the first. Does the crucial difference relate to UG?

One possible reason for this difference is the very existence of the L1 in the mind. L2 learners already know an L1; they have incorporated UG into an S_s. The initial L1 state in the child's mind has no language-specific knowledge; the initial state of the L2 learner already contains one grammar, complete with principles and actual parameter settings. With different starting points for L1 and L2 acquisition, it would hardly be surprising that the end result were different. The L2 learner's mind already contains an L1 grammar, with which the L2 has to coexist.

The final product of L1 acquisition is linguistic competence, which is complete by definition: a native speaker's competence is whatever a native speaker knows, neither more nor less. But the final point in L2 learning is hard to define: what *is* a normal L2 speaker? A person who can effortlessly pass for a native speaker in all circumstances? A person who can barely manage to order a coffee in a restaurant? A person who can translate Shakespeare? A person who can interpret the small print in a contract? One possibility is to identify the final grammar of L2 learners with that of adult L1 monolinguals: L2 acquisition is complete when L2 learners have the same knowledge of an L2 as monolinguals do of a first. Chomsky himself argues for the 'common-sense' view that only the complete adult knowledge of language counts, rather than intermediate states:

> We do not for example say that the person has a perfect knowledge of some language L similar to English but still different from it. What we say is that the child or foreigner has a 'partial knowledge of English' or is 'on his or her way' towards acquiring knowledge of English, and if they reach this goal, they will then know English. (Chomsky, 1986a, p. 16)

But such 'ambilinguals' (Halliday et al., 1964) or 'balanced bilinguals' functioning equally in both languages are rare, at best a small and untypical minority of L2 speakers. Most people are substantially less efficient in their L2 than in their first; many learn little of the L2, sometimes despite their best efforts. At any moment 80 per cent of the L2 learners of English in the world are supposed to be beginners, implying that few go on to become intermediate or advanced speakers. If the final goal of L2 learning is the knowledge of the native adult monolingual, paradoxically very few L2 learners actually reach it. While L1 competence is whatever it is, L2 competence is usually defined as what is *not*, in short as if it were L1 competence.

6.5.2 *The states metaphor*

The other powerful metaphor for language acquisition introduced in chapter 2 was to see it as the development of the language faculty in the mind from a zero state S_0 of knowing no language to a steady-state S_s of knowing all the grammar of the language: 'the language faculty just has states: one state is the initial state; others are the stable states that people reach somehow, and then there are all kinds of states in between, which are also real states, just other languages' (Chomsky, 2000a, p. 131). A version of this states metaphor for SLA is seen in figure 6.4, using S_i (initial) and S_t (terminal to distinguish the L2 process from the L1 S_0 and S_s (Cook, 1985).

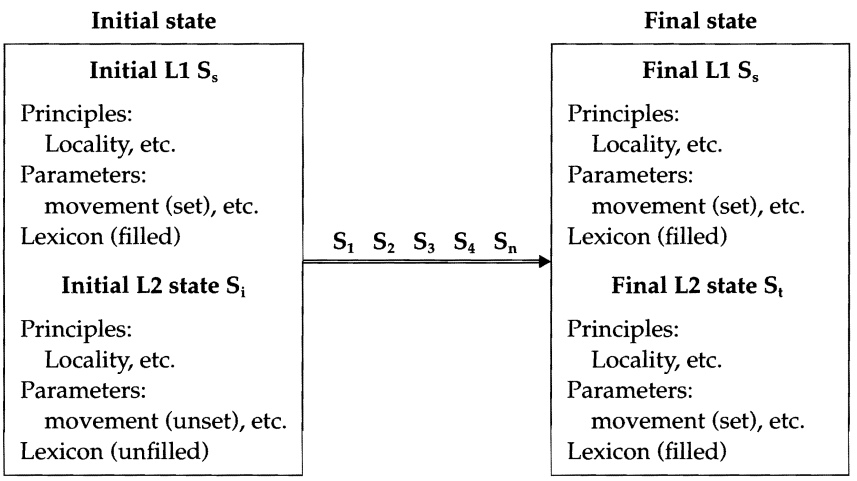


Figure 6.4 The development of the language faculty in second language acquisition: initial to final states

Let us talk through the implications of this states metaphor for SLA:

- the **initial state**. In L1 acquisition the initial zero state S_0 is simply UG. The distinctiveness of L2 acquisition is that one language has already been acquired. In other words the L2 initial state S_i differs by definition in already

containing a steady-state S_s grammar of some human language (setting aside early childhood simultaneous bilingualism). The principles have been instantiated, the lexical items with their parameters have been acquired. A crucial issue is then the relationship between the components of the initial L2 state S_i: does the L2 grammar start as a blank form with none of the elements filled in? Or does it start with the contents of the L1 grammar, which have to be adapted into an L2 form? Or does the S_i have a single grammar rather than two separate grammars in the form of an L1 steady state plus an L2 zero state?

- the **final state**. The final state might alternatively consist of two grammars – the S_s of the L1 and the S_t of the L2. One reason for calling the final L2 state S_t rather than S_s is indeed the doubt whether L2 learning leads to a steady state constant across learners. Unlike L1 acquisition, where essentially all children acquire the same linguistic competence, L2 learners reach differing levels in the L2, some able to function at high levels barely distinguishable from native speakers, others at a minimal level. One issue is 'ultimate attainment': do L2 learners ever reach a final state equivalent to a monolingual S_s? Again, the overall issue is the relationship between these two grammars; are they indeed entirely separate? Do they influence each other so that the L1 S_s is affected by the L2? Is there in effect one grammar for the two languages?
- the **intermediate states and processes**. In between the initial L2 state S_i and the final L2 state S_t comes a range of intermediate states S_1 to S_n, say. The questions are how these intermediate states relate to UG and what pushes the learner to move from one state to another, say in the input.

One controversy in L1 acquisition essentially concerned whether the child started with a complete grammar or a partial grammar, as we saw in chapter 5. The parallel issue in L2 acquisition is whether the learner starts by building the L2 from the L1 grammar, develops the L2 grammar separately or somehow merges the two.

6.6 Hypotheses of the initial second language state

In the 1980s the role of UG in L2 learning was expressed as a metaphor of 'access' to UG. Cook (1985) put it as a choice between three possibilities:

- L2 learners start from scratch; they have **direct access** to UG and are uninfluenced by the L1.
- L2 learners start from their knowledge of the L1; they have **indirect access** to UG via the L1.
- L2 learners do not treat the L2 as a language at all; they have **no access** to UG and learn the L2 without its help.

The concept of access, however, did not mesh very well with the states metaphor since it implied UG was something independent of the grammar on which the learner could draw rather than a state of the language faculty (Cook,

1994a). The SLA research of the 1990s moved on to conceptualizations of SLA based on the states metaphor and spent much effort on debating the possible initial states in L2 acquisition. Given that the mind contains a steady-state S_s of the language faculty for the L1 (setting to one side child L2 learners whose L1 competence may still not be adult-like), where is the space for the knowledge of the new language? Some of the alternatives for the initial state in L2 acquisition that were entertained were:

(a) the L2 learner has no UG to build on
(b) the L2 learner has a second copy of UG to build on
(c) the L2 learner can build on UG as incorporated in the S_s
(d) the L2 learner can partly build on UG.

These alternatives are discussed below.

6.6.1 *The second language learner has no Universal Grammar to build on*

UG has in some way been used up in L1 acquisition: the principles have been instantiated, the parameters set, the vocabulary learnt: the initial L2 state does not contain UG. Since UG is unavailable, L2 grammars must come from some other source than UG and be learnt in some other manner, say by using general learning processes. This is then the **No UG Hypothesis**.

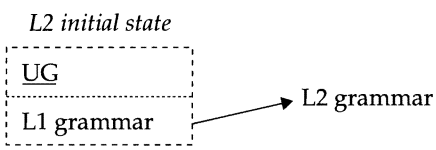


Figure 6.5 Initial second language state: No Universal Grammar Hypothesis

Several types of evidence are claimed to demonstrate UG is not used in L2 learning. In German there is a tension between the underlying order of the sentence, which is SOV, and the usual 'canonical' order in declarative sentences, which is SVO. Clahsen and Muysken (1986) found that L1 learners started with the correct underlying SOV order even if this yielded sentences that were wrong;

(4) Ich shaufel haben.
(I shovel have)
I have a shovel. (SOV)

but that L2 learners started with the correct surface SVO order even if they were wrong in the underlying structure.

(5) Ein herr verkaufen blumen.
(a master sells flowers)
A master sell flowers. (SVO)

The L2 sequence was claimed to demonstrate lack of access to UG: 'by fixing on an initial assumption of SVO order, and then elaborating a series of complicated rules to patch up this hypothesis when confronted with conflicting data, the L2 learners are not only creating a rule system which is far more complicated than the native system, but also one which is not definable in linguistic theory' (Clahsen and Muysken, 1986, p. 116).

In addition several general arguments in favour of the total detachment of UG from L2 acquisition have been put forward:

- the knowledge of the L2 is not so complete as that of the first in monolinguals (Schachter, 1988; Bley-Vroman, 1989);
- some L2s are more difficult to learn than others, for instance Chinese versus Italian for speakers of English (Schachter, 1988);
- the L2 gets fossilized rather than progressing inevitably to full native competence, as does the L1 (Schachter, 1988);
- L2 learners vary in ways that L1 learners do not, few achieving proficiency in their L2 comparable to that in their first.

Those who claim UG has no place in the initial state have therefore sought to explain how an L2 can be learnt without UG. The typical solution is to invoke general problem-solving combined with the knowledge of the L1 (Bley-Vroman, 1989), i.e. what was earlier called 'other faculties'.

However, it should be noted that those who say UG is not involved in L2 acquisition do not usually oppose its existence in L1 acquisition, though this is denied by many other acquisition researchers. To apply the UG model to SLA at all involves accepting that it *is* relevant to L1 acquisition. As White (2003, p. 60) puts it: 'All the initial-state proposals . . . presuppose the following: UG is constant . . . ; UG is distinct from the learner's L1 grammar; UG constrains the L2 learner's interlanguage grammar.' In a sense this position commits one to the growth view of L1 development outlined in chapter 4, since the UG in the child's mind is no longer in its initial untouched state available for L2 acquisition but has been changed in some ways by the acquisition of the L1.

6.6.2 *The second language learner has a second copy of Universal Grammar to build on*

This second copy is distinct from the S_s. The learners can build an L2 grammar from scratch in exactly the way they did a first, instantiating the principles and setting the lexical parameters: the initial state contains a new example of UG alongside the S_s from the acquisition of the L1. This is a version of the age-old idea that SLA mirrors L1 acquisition, couched in terms of UG. It implies that the initial state can contain a copy of UG which has remained unaffected by the acquisition of an L1; presumably each time a new language is encountered the same happens again. This is the **Full Access Hypothesis**, which in its purest form sees L2 acquisition as uninfluenced by the L1 that the learner has already acquired.

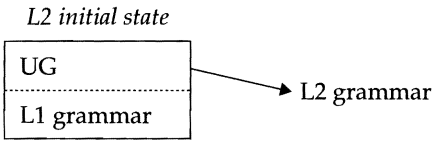


Figure 6.6 Initial second language state: Full Access Hypothesis

One way to demonstrate this is to show that L2 learners indeed know principles and parameters they could not have learnt. A test-case can be whether L2 learners of English who do not have movement in their L1 display restrictions on movement in the L2, first tested by Naoi (1989) with Japanese learners of English.

On the basis of an early claim of Chomsky's (1971) that all linguistic processes are dependent on the details of the structure to which they apply, known as the principle of structure-dependency, Cook (2003) tested the grammaticality judgements of 140 L2 learners of English with six different L1s. The sentences tested were either violations of movement restrictions such as:

(6) *Is Joe is the dog that black?

or were unmoved forms such as the relative clause:

(7) Joe is the dog that is black.

or questions with relative clauses:

(8) Is Joe the dog that is black?

Overall 131 out of the 140 subjects correctly rejected the sentences with violations, i.e. above 83 per cent, though they scored worse on the other sentence types. The subjects included people whose L1s need structure-dependency for movement (Polish, Dutch, Finnish) and whose L1s did not have overt movement (Japanese, Chinese, Arabic). Japanese learners scored 95.1 per cent correct, Chinese students 86.7 per cent, Arabic 87.1 per cent, and Polish, Dutch and Finnish over 99 per cent. So the knowledge about structure-dependency in questions did not come from the S_s in their minds. Nor is it likely that any other sources of evidence were available to them from teachers' explanations, etc.

It is perfectly possible to argue that structure-dependency is not the correct analysis of these sentences and the experiment therefore misses the point; indeed chapter 1 analysed them in terms of Locality. But, unless there is some means by which the learners have acquired this restriction from the environment, it must have come from within their own minds.

More generally, the strong similarities in L2 acquisition between learners with different L1s have been a constant theme in SLA research, ranging from the common order for the acquisition of grammatical morphemes of the 1970s (Dulay and Burt, 1973) to the Basic Grammar shared by speakers of six different languages learning five L2s of the 1990s (Klein and Perdue, 1997). If L2 learners are all rather similar, UG cannot differ much between them. To return to the experiment, while

there were differences between the speakers of L1s with movement and without movement, these relate to levels of success, not to failure to detect the violations. At best the L1 S_s makes a small difference to the use of UG in the acquisition of an L2.

The chief proponents of Full Access are Epstein et al. (1996), who claim that UG is accessible throughout L2 acquisition from the initial stage on; the starting point in L2 acquisition is the UG in the mind rather than the L1. This then separates UG from the L1 grammar and sees it as 'continuously available to assist in the construction of various language-specific grammars' (Flynn and Lust, 2002, p. 98). Hence there is not so much an L2 initial state as an initial state for acquisition of *any* language, whether L1, L2 or L3. There should therefore be a strong similarity between L1 and L2 acquisition (a large area of L2 acquisition research debate in the 1970s, summarized in Cook et al. (1979)). Flynn and Lust (2002) support this Strong Continuity Hypothesis with evidence that both Japanese and Spanish learners of L2 English are governed by the UG principles and parameters of word order, not by their L1. The differences between them, whatever the cause, are not due to L1 transfer.

6.6.3 *The second language learner can build on Universal Grammar as incorporated in the S_s*

In this case, the elements of UG are still present in as much as they are reflected in the L1 S_S. They can be adapted and reset to create a second grammar for the L2, using the still accessible UG in the mind; those aspects that are not already instantiated in the L1 S_S are still there in the latent UG to be used if needed. The starting L2 grammar is effectively a clone of the S_S, and is built up gradually, using the UG principles and parameters, into a distinctive grammar of its own. This is known as the **Full Transfer/Full Access Hypothesis**: 'the initial state of L2 acquisition is the final state of L1 acquisition' (Schwartz and Sprouse, 1996, p. 41).

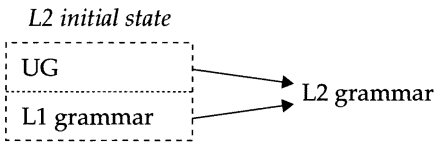


Figure 6.7 Initial second language state: Full Transfer/Full Access Hypothesis

This hypothesis requires evidence of two types: one that the L1 is indeed fully present in the L2 grammar from the beginning; the other that the developing grammar makes unfettered use of UG. Combining these two claims is then tricky. The original manifesto in Schwartz and Sprouse (1996) relied on data from a single Turkish L2 learner of German, which were argued to show use of L1 Turkish, for example in the position of verbs in German subordinate clauses, but a progressive restructuring of the grammar according to overall options of UG such as the Case Filter. In a sense this qualifies the Full Access Hypothesis shown above, in which UG was seen as fully accessible, by stressing that, in L2 learning, another language is already present in the initial state.

6.6.4 *The second language learner can partly build on Universal Grammar*

In this view the elements of UG are not available in their entirety. The initial state has a defective clone of UG present. The various alternatives can collectively be called the **Partial Access Hypothesis**.

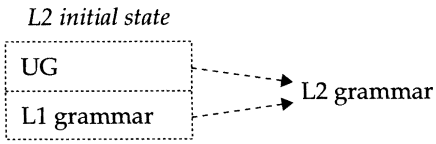


Figure 6.8 Initial second language state: Partial Access Hypothesis

Among the different alternatives for partial access that have been suggested are the following:

1 The initial L2 state consists of an L1 grammar and UG, but, while the grammar contains lexical categories such as NP and PP, it lacks functional categories such as IP and DP. This is known as the **Minimal Trees Hypothesis**: 'only *lexical categories* are present at the earliest stage of both L1 and L2 acquisition and . . . during acquisition *functional projections* develop in succession' (Vainikka and Young-Scholten, 1996, p. 7). It resembles the account of L1 acquisition called the Gradual Development Hypothesis in chapter 5; not just L1 children but also L2 learners start with defective grammars lacking the vital element of functional categories. The hypothesis depends on the missing elements in the sentence appearing in a clear stagiation. The developmental evidence adduced by Vainikka and Young-Scholten (1996) comes from L2 learners of German. At the first stage they transfer the L1 word order to the 'bare' VP without any inflections:

(9) Oya Zigarette trinken.
Oya smokes cigarettes. (literally 'drinks')

At the second stage they use an IP with an initial head:

(10) Ich sehen Schleir.
I see the veil.

At the third an Agreement Phrase (AgrP) with agreement of auxiliaries

(11) Ich kaufe dich Eis.
I will buy you an ice-cream.

Thus the learners' repertoire of phrases expands from the lexical VP to the functional IP and AgrP. Again this is based on a small amount of production data, coming from five L2 learners, one Spanish and four Italian for the developmental data.

To some extent this narrows down the type of explanation that can be offered. In L1 acquisition, the reason for missing functional categories might be purely linguistic properties of the language faculty that need to develop in sequence or it might be the actual maturation of the child. In SLA, the maturational explanation is ruled out as the learners are either fully mature or at any rate past the usual L1 stages (with the exception of simultaneous early bilingualism), leaving only the internal linguistic explanation. Thus SLA research can feed back into general issues of language acquisition by ruling out some of the alternative explanations for language development (Cook, 1981).

2 The initial L2 state consists of an L1 grammar and UG; the L1, however, has the functional categories already present but minus their parameter settings. This is known as the **Valueless Features Hypothesis**; 'the very act of transfer obliterates the values associated with features located under functional heads' (Eubank, 1996, p. 73). As evidence Eubank cites Wode's studies of his German-speaking children acquiring English negation (Wode, 1981).

3 The initial state has the L1 grammar minus the unset parameter-settings; as learners progress, they fill in the parameter values for the L2. This is known as the **Failed Functional Features Hypothesis** (Hawkins and Chen, 1997); 'Beyond the critical period the virtual unspecified features disappear, leaving only those features encoded in the lexical entries for particular lexical items' (Hawkins and Chen, 1997, p. 216). Access to UG is then through the settings of the L1 grammar. Their evidence for this is the gaps that Chinese learners have in their knowledge of English relative clauses because they see the relative clause as a structure of its own rather than one derived by movement – Chinese lacks relative clause movement while French does not. One crucial area they tested was resumptive pronouns, as seen in:

(12) The man who he lives next door has left.

Only 38 per cent of Chinese elementary learners rejected this compared with 81 per cent of French learners; at intermediate level this rose to 55 per cent for Chinese, 90 per cent for French; and at advanced level to 90 per cent and 96 per cent. The Chinese learners were starting at a disadvantage because they did not have the features available for specifying movement while the French learners did.

EXERCISE 6.2

Here is a sample of written L2 learner English, typical of an early stage of acquisition. Can you find in it any evidence for or against the different hypotheses? If so, which does it confirm? If not, how valid can these hypotheses be if they do not cover features manifest in such a sample?

I ferom Israel. I em e merid woman I got one cheild 5 yeres oud I nov my eghit it na very good we live in egland about tou yers I live in Standford Hill London N16. The neme of my douther is Lee I love ther very mach I gout i sister in Israel and al my famili I be in the harmy en it uous wunderfent last wik we went tu paris.

Hypothesis	L1 S_s involved	UG involved	Initial state consists of:
1 No UG access	✓	✗	L1
2 Full access	✗	✓	All UG
3 Full transfer/full access	✓	✓	UG plus L1 S_s
4 Partial access:			
(a) Minimal trees	✗	✓	L1 minus functional categories
(b) Valueless features	✗	✓	L1 minus features strength
(c) Failed functional features	✓	✗	UG principles and L1 minus parameter settings

Figure 6.9 Alternative hypotheses of the initial second language state

6.7 The final state of second language acquisition

At the other end of the states model of acquisition comes the final steady-state S_s, where acquisition is effectively complete. As we have seen, many researchers identify the final or ultimate state S_t with the monolingual native speaker, put by Spolsky (1989, p. 35) as 'Condition 2. . . . second learner language approximates native speaker language', but implicit in many remarks such as 'Unfortunately, language mastery is not often the outcome of SLA' (Larsen-Freeman and Long, 1991, p. 153) or 'failure to acquire the target language grammar is typical' (Birdsong, 1992, p. 706). The main issue of SLA research is seen as explaining this failure: 'children generally achieve full competence (in any language they are exposed to) whereas adults usually fail to become native speakers' (Felix, 1987, p. 140). The question of access to UG is largely couched in these terms of failure: access is demonstrated if the L2 learner's knowledge is identical with that of the native speaker. It might be, however, that the L2 learner's grammar is still a possible human language within the remit of UG principles and parameters but different from either the L1 or the language that the L2 learner is acquiring. UG is still available in L2 acquisition even if it results in a different grammar from the language the learners start from or the one they are learning.

This area of SLA research has then developed into disputes about rival hypotheses, mostly set up to explain the apparently 'missing' elements of the L2 learners' grammar; a lucid and balanced account of these is provided in White (2003). What started as a simple question about the involvement of UG in L2 acquisition has turned into a morass of sub-questions about technical details of the syntax, supported by interpretation of evidence from very small numbers of cases, usually reanalysing old production data. The papers dealing with the initial L2 state rely infuriatingly on the reader having a total command of the syntactic theories current at the moment of writing, as if the authors were writing syntactic theory rather than SLA research; to do real justice to the Eubank (1996) paper, for example, means having a working knowledge of the 1980s

syntactic work by Pollock (1989) and 1990s work by Chomsky (1995a) and Kayne (1994) inter alia, all major theoretical ideas, not to mention four at the time unpublished manuscripts; given that each of the theories that is drawn upon is one of many possible alternatives, all of which have a short sell-by date, it is hard to know the validity of this pot-pourri approach. Indeed the above presentation of the hypotheses skimmed over many of the details partly because of the complexity of the syntactic issues involved, partly because they would involve elaboration of syntactic areas or theories no longer current. Curiously enough the L2 acquisition research is far more difficult to get into than the L1 acquisition research.

If the UG approach is to make a real contribution to SLA research rather than becoming a ghetto of its own making, it needs to settle these claims with some more conclusive research and to proceed to more interesting questions, in particular broadening to deal with the whole of L2 acquisition rather than a narrow area of syntax based on missing categories and movement. Rather than using the richness of the UG Theory, this area seems to have become the kind of discussion associated with the question, attributed, probably spuriously, to St Aquinas, of how many angels can dance on the head of a pin.

From the perspective of multi-competence, the initial and the final L2 states are both states of a single mind, one containing knowledge of one language, the other knowledge of two or more. If any mind has the potential to learn more than one language, the principles and parameters of UG have to be established from people who know more than one language. Yet the UG-related research tests L2 learners against the UG established from monolinguals, rather than establishing UG from multilinguals using a poverty-of-the-stimulus argument and then seeing how monolinguals *fail* to acquire this 'multilingual-derived' UG. L2 users are not failures for not knowing a principle, such as the θ-Criterion, known by L1 speakers; the description of the L1 speaker's θ-Criterion is faulty if it cannot accommodate L2 users' grammars.

In particular, multi-competence has raised the issue of how the L2 knowledge in the S_t affects the other component of the final state – the L1 (Cook, 2003): French speakers who know English react against French sentences using the middle voice

(13) Un tricot de laine se lave à l'eau froide.
*A wool sweater washes in cold water.

compared to those who don't know English (Balcom, 2003); Japanese, Greek and Spanish speakers of English prefer the first noun to be the subject of the sentence in:

(14) The dog pats the tree.

more than do those who do not know English (Cook et al., 2003). Multi-competence, far from denying the existence of UG, insists that its description be based on the final language state of the normal human being, which includes more than one language.

By and large the Chomskyan approach has also had little effect upon language teaching. Chomsky's general ideas were used as a club to squash the behaviouristic audiolingual method at various times; 'these principles are not merely inadequate but probably misconceived' (Chomsky, 1966b, p. 153). For a generation, language teaching has, however, been concerned with E-language and pragmatic competence rather than with I-language and linguistic competence: 'communication' has been the key-word, followed by 'task'. Hence a theory based solidly on I-language, concerned primarily with syntax and disdaining communication as a derived function of language has had little to offer.

Some possible overall implications for language teaching follow:

- Teachers should concentrate on the aspects of syntax that will not be automatically acquired by the students from ordinary language input (Cook, 2001). Thinking of the language of the classroom as providing input for setting parameters may be a helpful addition to other ways of thinking about language teaching.
- It is useful to remember that language learning is not just learning how to behave in a range of situations, it is also the acquisition of knowledge in the mind. Teaching has tended to go overboard for a one-sided view of language; UG reminds us that language is mental knowledge as well as behaviour.
- Some of the descriptive devices used by UG Theory may be of use in explaining grammar to L2 students, say the pro-drop parameter. Mostly language teaching uses a mixture of types of grammatical description, as seen in Nassaji and Fotos (2004), ranging from traditional grammar to grammatical morphemes, but so far has not visibly incorporated such a usable aspect as pro-drop some twenty-five years after its beginning.

Further discussion can be found in Cook (1994b). The applications of multicompetence have been written up extensively elsewhere (e.g. Cook, 2002), mostly involving a denial of the native speaker as the chief target of language teaching, an assertion that the L2 has to be reconciled with the L1 in the learner's mind, thus bringing the L1 back into the classroom, and the realization that language teaching is in the business of teaching students knowledge, not how to behave.

EXERCISE 6.3 Applications Assuming that it is necessary for the following aspects of Government/Binding syntax to be taught in some way to L2 learners of English, try to devise ways of teaching them:

- how to use *himself* versus *him*
- how to have subjects in all sentences
- how Verbs come before objects, subjects before Verbs.

Discussion topics

1 Do you think that people can attain the same end-state in SLA as in L1 acquisition?
2 What would you consider the best of the competing hypotheses for the initial state of SLA? Why?
3 Does the poverty-of-the-stimulus argument really apply to SLA?
4 To what extent can 'other faculties' be involved in SLA that are not involved in L1 acquisition, as seen for instance in the use of grammatical explanation in L2 teaching?
5 In the light of the discussion here, to what extent does SLA resemble L1 acquisition? Or is it really a distinct process?
6 Do you agree with the implicit assumption in much UG research that the 'normal' state of the language faculty is that of a monolingual rather than a bilingual?
7 To what extent can or should UG research in SLA draw on the concepts and research methodologies of L1 acquisition?

7 Structure in the Minimalist Program

This chapter and the one that follows introduce the syntactic ideas that Chomsky has been proposing since the early 1990s, known as the Minimalist Program (MP). This developed out of the Principles and Parameters (P&P) theories of the 1980s in much the same way that these theories developed out of work done in the 1960s and 1970s. In some respects, the MP diverges more radically from earlier assumptions, abandoning much of what was once held to be central. In other respects, however, it maintains some of the core assumptions that separate P&P Theory from its predecessors, particularly the notions of universal linguistic principles and language-particular parameters; hence it is equally entitled to be included under the rubric of P&P Theory.

The MP is not a unified theory of language as Government/Binding (GB) Theory could claim to be. To start with, it is called a 'programme' not a theory. This is intended to convey a certain degree of tentativeness about its leading ideas, which are based on relatively new concepts whose properties are not yet fully understood and are very much open to programmatic exploration. Indeed, Chomsky readily admits that many questions posed by the MP are poorly understood and perhaps even premature:

> Questions of this kind are not often studied, and might not be appropriate at the current level of understanding, which is, after all, still quite thin in a young and rapidly changing approach to the study of a central component of the human brain, perhaps the most complex object in the world, and not well understood beyond its most elementary properties. (Chomsky, 2000c, p. 94)

Given this tentativeness, it is not surprising that there have been several radical changes to Chomsky's thinking within the MP itself. The early ideas up to 1995 as presented in 'A Minimalist Program for linguistic theory' (Chomsky, 1993) are distinct from those presented in the seminal chapter 4 of Chomsky's book *The Minimalist Program* (1995a) and different again from his later writings, particularly 'Minimalist inquiries' (2000c), 'Derivation by phase' (2001a), which has given rise to what has come to be known as 'Phase Theory', 'Beyond explanatory adequacy' (2001b) and 'On phases' (2005c).

This chapter does not treat the MP historically, nor does it cover many of the technical details found in Chomsky's own writing. Instead it outlines the main flow of Chomsky's recent work, to give some insight into his current thinking and to allow the reader an idea of where the programme is heading. After a brief introduction to the central concepts of the MP, we will concentrate on issues of basic structure and how these are handled within the framework. The next chapter shifts attention to issues concerning movement. In other words the division between these two chapters mirrors that used for GB Theory in chapters 3 and 4.

7.1 From Government/Binding to the Minimalist Program

Before presenting the ideas of the MP, it is useful to consider why there was a need for it in the first place. It is not that the MP developed out of the failure of GB Theory, for GB was, and still is, very successful, providing us with many novel insights into syntactic phenomena, with a breadth of description and a level of explanation not previously attained. Indeed, there are still syntacticians whose current work would be best categorized as GB syntax, even though some wave hands in the direction of the MP. The same is true of researchers into first language (L1) acquisition where, despite some overt claims to be minimalist, the bulk of work is carried out within a GB framework. It is because of the very success of GB that this theory takes up over half of this book, as we believe it still provides the non-expert with the best introduction to modern syntax and that GB Theory is essential to an understanding of all subsequent developments. This is not to say that GB Theory was perfect, but problems which face a theory are usually an indication for its development rather than abandonment; hence it is not true to say that the MP arose from the ashes of a failed GB.

But GB was nevertheless not perfect – the human language faculty is far too complicated for us to understand it fully at such an early point in our scientific investigations into it – and GB did suffer from certain conceptual and empirical problems, as seen in earlier chapters. Moreover, the attempt to solve certain of these problems led to the undermining of previously well-established concepts, which in turn gave rise to new problems. GB's 'rich deductive' nature in a sense made this inevitable: as all GB modules interact with others in complex ways, tinkering with one part of the theory naturally had huge consequences for other parts. This was obviously an advantage in accounting for how languages could seem vastly different from each other at the surface and yet be distinguished by slight changes in parameter settings at the deeper level of Universal Grammar (UG), at which language acquirers operate, but it undoubtedly presented problems for the refinement of the theory itself.

To take an example, the notion of an A-position was well defined in early GB Theory as a potential θ-position – positions to which θ-roles could potentially be assigned, such as subject or object positions, but not adjunctions or the specifier of CP (see chapter 3). However, the development of the VP-Internal Subject Hypothesis gave rise to two distinct subject positions: the specifier of VP and the specifier of IP. The specifier of VP is obviously an A-position as it is a θ-position.

However, it is no longer obvious that the specifier of the IP is an A-position as it is clearly not even a potential θ-position under these assumptions. In fact, the specifier of IP has much in common with the specifier of CP: both are specifiers of a functional head and both serve as the landing site for certain phrasal (XP) movements (X stands for any a category, i.e. N, V etc.).

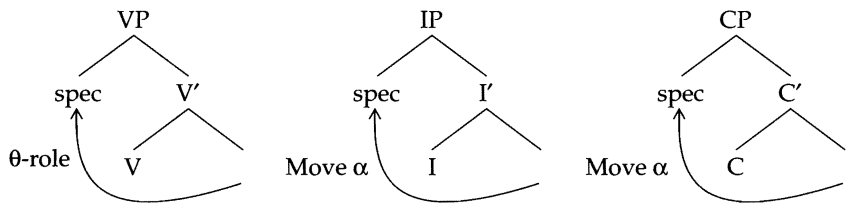


Figure 7.1 Similarities and differences between VP, IP and CP specifiers

Yet the specifier of CP is an Ā-position. Despite this, there is still the need to consider the specifier of IP as an A-position rather than an Ā-position, as movement to spec IP (A-movement) differs from movement to spec CP (Ā-movement), and binding from spec IP (A-binding) differs from binding from spec CP (Ā-binding). But to group spec VP and spec IP together as A-positions when these positions have very little in common, and to separate spec IP from spec CP when these have much in common, comes across as stipulatory and counter-intuitive.

Issues such as these might have been addressed within GB, however, if it were not for the fact that at the same time as they began to assert themselves there were other changes beginning to take shape which suggested an entirely different approach. We will introduce these in the next section.

7.1.1 *The development of the concept of economy*

The MP did not just materialize out of thin air. Although it rejects many of the assumptions made in GB, it is a development from ideas formed within this earlier framework. The notion of economy that Chomsky developed in the late 1980s within the GB framework had a direct bearing on the development of the MP, at least in its early stages.

The seed of these ideas can be traced back to Chomsky's 1986 introduction of the Principle of Full Interpretation (Chomsky, 1986a): every element in a structure must be interpreted in some way, i.e. there are no superfluous elements in the structure of language. This was originally intended to set the Theta Criterion on a deeper, more explanatory footing. The Theta Criterion requires arguments and θ-roles to be in a one-to-one correspondence so that each argument bears exactly one θ-role and each θ-role is assigned to exactly one argument, as seen in chapter 3. Specifically, no more arguments can occur than are required to bear θ-roles, and no more θ-roles can occur than are required to be assigned to arguments:

(1) a. * John loves Mary Susan. (too many arguments)
b. * John requires. (too many θ-roles)

Another way to put this is that the linguistic system does not allow superfluous arguments or θ-roles. When looked at in this way, it can be seen that this requirement is more general still: the linguistic system does not allow superfluous elements of any kind. This is readily apparent if we compare natural language to mathematical or logic systems. In these, the syntax is kept as simple as possible, as it is the 'meaning' which is obviously important. So long as the mathematics is correct, it wouldn't matter whether a mathematical sentence is represented as $2 + 2 = 4$, or as $+(2, 2) = 4$. For this reason, there are many mathematical sentences which are allowed by its simple grammar which appear to make no sense whatsoever. Consider the following:

(2) a. For all numbers n, $n + 1 > n$.
b. For all numbers n, $2 + 2 = 4$.

Although not very interesting, sentence (2a) happens to be a true statement: adding 1 to any number makes it larger than the original number. Moreover (2a) is grammatical. It is made up of a quantifier expression (*for all numbers n*) followed by something that would be a grammatical sentence on its own ($n + 1 > n$), though because it contains an unknown element n it could not be evaluated. A simple rewrite rule of the grammar of this 'language' might therefore be:

(3) $S \rightarrow Q\ S$

This rule says that a grammatical sentence can be formed by placing a quantificational expression (such as *for all numbers n*) in front of another sentence. The important point is that (2b) also conforms to this rule and so it is also grammatical. The difference between (2a) and (2b) is the fact that in the (2a) example, the quantifier expression performs a meaningful role as there is an unknown 'variable' number (n) in the following sentence that the quantifier tells us how to interpret: it says that if we replace n in this sentence with any number the result will be true. In (2b), however, there is no such variable in the sentence and hence the quantifier expression serves no purpose. In mathematics this is not a problem: we ignore the quantifier and interpret the sentence as though it were not there; it's simply superfluous to the meaning of the sentence. If we wanted to exclude sentence (2b) as ungrammatical, a far more complex grammar would be needed, as the rule in (3) is clearly inadequate. But, given that in mathematical languages, grammatical issues such as these are unimportant, there would be no point in complicating the grammar.

Consider now the following sentence:

(4) * Every the cow chewed the cud.

Clearly this is ungrammatical in English. Yet it is identical in all relevant respects to the mathematical sentence in (2b): there is a quantifier at the beginning of the sentence which serves no purpose. The fact that this sentence is ungrammatical demonstrates that natural languages are fundamentally different from mathematical languages in this respect. In natural languages we can't just ignore superfluous

elements in a sentence and pretend they are not there. In this respect natural language grammars seem more complex than mathematical languages.

Observations such as these led Chomsky (1986a) to propose a basic principle of natural languages:

(5) **Full Interpretation**
Every element in a sentence must be given an interpretation.

It is easy to see how the Theta Criterion falls out from this principle: θ-roles are interpreted on the arguments that bear them and arguments are interpreted with respect to the θ-roles that they bear. It therefore follows that all theta roles must be assigned to an argument and all arguments must be assigned a θ-role, exactly what the Theta Criterion stipulates.

The discovery of this basic principle sowed in Chomsky's mind the seed of a characterization of language that was to emerge some years later under the notion of **economy**. The idea is that although principles like Full Interpretation complicate the grammar, their effect is to make language economical: everything that appears in a sentence serves a purpose. It is highly uneconomical to allow superfluous elements in a sentence which are effectively ignored as this creates the situation in which a language might have an infinite number of expressions which amount to the same sentence once all superfluous elements are discounted.

Full Interpretation was the foundation of what Chomsky called **economy of representation**: 'there can be no superfluous symbols in representations. This is the intuitive content of the notion of Full Interpretation' (Chomsky, 1986a, p. 151). At the same time, he also noted other economical aspects to language. Chapter 4 discussed facts about head movements in English such as that main verbs do not move out of the VP, but auxiliary verbs do, hence yielding the contrasting positions of these verbs in the following sentences:

(6) a. He quickly *hid* the evidence.
b. He *had* quickly hidden the evidence.

For main verbs, it appears that the inflection moves from I to V, while auxiliaries move from V to I to become attached to the inflection. Pollock (1989) had suggested that main verbs cannot move out of VP in English because they are unable to assign their θ-roles from the inflectional position and hence a Theta Criterion violation would follow if they were to move. As auxiliary verbs do not assign θ-roles, they are not constrained in this way. However, note that, although I to V movement seems possible in English, it is impossible to move the inflection when there is an auxiliary:

(7) * He quickly had hidden the evidence.

Thus, when a verb *can* move, it *must* move: only if a verb cannot move will the inflection move. Chomsky's (1991a) account of these observations made use of the fact that I to V movement is a lowering movement, which should not be possible as the trace left behind would be higher than the moved element and as

such would not be governed, in violation of the Empty Category Principle (ECP). To circumnavigate the ECP, which applies at Logical Form (LF), Chomsky suggested that the overt lowering of the inflection must be accompanied by a covert raising of the verb to the inflection position to satisfy the ECP at LF. For reasons too technical to go into here, this covert raising overcame the difficulties that would prevent the verb from assigning θ-roles if it moved out of the VP. Thus, if lowering movements are always accompanied by a subsequent raising movement, it follows that an initial raising of the relevant element will always produce a shorter derivation as there will be no subsequent lowering accompanying it. Given that raising movements are obligatory (auxiliary verbs can and must raise), the system seems to prefer shorter derivations to longer ones. Longer derivations are only possible if forced (main verbs cannot move out of VP, so inflection lowering followed by covert raising is allowed). The preference for shorter derivations Chomsky called **economy of derivation**.

Another aspect of the economy of derivation, Chomsky claims, is that some processes are preferred over (are 'cheaper than') others. In his paper on verb movement (Chomsky, 1993), Chomsky also claimed that '*do*-insertion' was used as a means to support an unbound tense morpheme only as a last resort because it makes use of a language-specific device: the insertion of the English dummy auxiliary *do*. Movement, on the other hand, is a part of UG and so 'comes for free'. Therefore, if movement of either the verb or the inflection is possible, movement will take place and only when neither is possible will *do*-insertion be resorted to. The idea that some processes 'come for free' while others bear a cost is a vital part of Chomsky's current thinking.

A further aspect of the economy of derivation can be seen from the perspective of Relativized Minimality, described in chapter 4. Recall that this principle claims that all movements must be to the nearest possible position, where the range of possible positions is determined by the properties of the moved element. Simply put, this principle favours shorter movements over longer ones. It is easy to see how this fits with the idea of economy: the more distance covered by a movement, the costlier it is, and hence there is pressure to keep the links between the elements in a movement chain to a minimum. We might therefore call this principle the **Minimal Link Condition**. Of course, the Minimal Link Condition seems to be in conflict with the requirement that derivations should contain the fewest steps: the shorter the movements, the more movements needed to cover the distance. One possible solution to this conflict might be that chains, i.e. the moved element and all its traces, are added to a structure as a single element in one derivational step. This would allow the Minimal Link Condition to be upheld without lengthening the derivation.

A grammar which determines grammaticality in terms of principles such as 'shortest derivation' or 'minimal links' is very different from the standard GB way of looking at things. In standard GB Theory, any derivation would produce a grammatical expression, providing it did not violate a constraint or produce a structure that was filtered out. In other words, the grammar could do anything that it could get away with. A more economical perspective, however, would suggest doing nothing unless forced to do otherwise. For example, in GB the principle Move α sanctioned the movement of any element to any position, unless a given

movement violated some other principle of the grammar. In the MP, the movement process operates only if it has to. Obviously, this leads to a very different kind of explanation for certain linguistic facts. Consider the following:

(8) * John$_i$ seems [t$_i$ is tall].

GB explains the inability of the subject to move out of the embedded finite clause in terms of the violation of the binding principles: the trace left behind by an A-movement such as this is an anaphor and anaphors must be bound in their governing categories (p. 167), in this case the embedded finite clause. An economic system, on the other hand might account for the ungrammaticality of (8) as due to the fact that there was no need for the subject to move: the subject already has its Case in the lower subject position and hence does not need to move to the higher subject position to get Case.

Perhaps the most important consequence for the MP was something that we may refer to as **economy of grammar**, though Chomsky never used this term himself. The core idea behind a minimalist approach is that analyses should proceed on the minimal number of assumptions and make use of the minimal number of grammatical mechanisms. To some extent this harks back to a very old idea concerning grammatical analysis: the notion of **elegance** has been used to favour one analysis over another for a long time. Although there is no reason to expect that the language faculty naturally developing alongside other biological organs would have to be elegant (functionality and robustness seem to be more favourable properties of other biological systems), the fact that a consideration of elegance has time and time again shown itself to lead to insight into the workings of language indicates that it is a central linguistic feature.

In more recent writing Chomsky has cast this issue in terms of 'language design' and, more specifically, in terms of the question of how 'perfect' language is, as briefly outlined in chapter 1: 'Suppose that a super-engineer were given design specifications for language: Here are the conditions that FL [the language faculty] must satisfy; your task is to design a device that satisfies these conditions in some optimal manner (the solution might not be unique). The question is: How close does language come to such optimal design?' (Chomsky, 2000c, p. 94). Of course, this raises the question of what the design specifications for language are, which Chomsky answers by drawing attention to the need for the language faculty to be accessible by the other faculties which make use of it. Figure 7.2 draws on chapter 1 to show the general model involved.

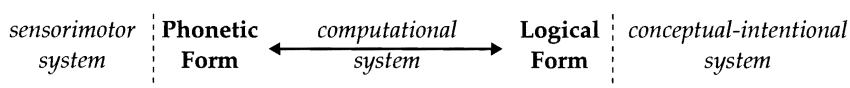


Figure 7.2 The computational system

The central computational system has to interface successfully with other faculties that impose their own restrictions on what goes on in the language faculty, known as **Bare Output Conditions**. Two faculties appear to make use of, or **interface with**, the language faculty: the sensorimotor systems for expressing and perceiving language and the 'systems of thought' for conceiving and understanding

linguistic expressions. The conditions that these impose are that the expressions formed by the language faculty must be 'interpretable', in the relevant sense, at the point of interface between them and the language faculty. In other words, at the point where the language faculty meets with these other faculties, the formed linguistic expression must meet a condition of Full Interpretation defined in a way relevant to the interpreting system – semantically interpretable for the conceptual-intentional system and phonetically interpretable for the sensorimotor system. In turn this implies that two interface levels of representation are required, as phonetic information is clearly not interpretable semantically and semantic information is not interpretable phonetically. In an optimum design, these two would be the only levels of representation required, obviously corresponding to what are called Phonological Form (PF) and Logical Form (LF) in GB, as described in chapter 3. We therefore see that the MP expects that internal levels of representation of the language faculty, such as D-structure and S-structure, lie outside the minimal requirements to meet design specifications and, if the language faculty is perfectly or near perfectly designed, they are surplus to requirements.

To summarize, the MP was born out of an emerging notion of economy, developed in GB Theory, which suggested that language was based on different kinds of principles from those which had previously been considered: an economic system would only operate principles if they were required. A notion of economy in the grammatical system itself also led Chomsky to consider questions of language design in general and more specifically the question of how 'perfect' the language faculty is. In an optimally designed system the only things that would be needed would be those things required by other systems that the language faculty interfaces with. This, then, is the essence of a minimalist grammar.

ECONOMY

Economy of representation: There should be no superfluous elements in the representation of a structure. The main principle which demonstrates this kind of economy is **Full Interpretation**, which states that every element in an expression must receive an interpretation.

Economy of derivation: There should be no superfluous steps in a derivation. This kind of economy prefers shorter derivations which make the fewest steps. Short moves are also preferred to long ones.

Economy of grammar: This is the heart of a minimalist approach to language. It investigates the possibility that human language design meets the Bare Output Conditions imposed on it from those systems with which it interfaces in an optimal way.

7.2 Basic minimalist concepts

As seen above, Bare Output Conditions demand the existence of two interface levels, LF and PF. For the time being, we will concentrate on how the LF representation is formed, returning to PF a little later.

The language faculty must contain the mechanisms which build LFs for each expression of the language. Clearly these mechanisms must have material to work with – the lexical items that form the basis of a structure. Chomsky envisages the process as follows. Starting with a set of lexical items, the derivation of an expression proceeds step by step. At each step some new part of the structure is built, the final step being the fully formed LF:

(9) lexical items
|
LF

Here the line between the lexical items and LF represents the step-by-step process of building the LF, referred to by Chomsky as the **computational procedure**. The minimal assumption about this procedure is that there is some operation which forms structures by simply combining any two elements it has to work with. Diagrammatically this process can be represented as follows:

(10) X Y ⟶ Z
X Y

All trees are then built cumulatively by combining two elements into one step by step. Chomsky terms this procedure **merger** and the operation which carries it out **Merge**: 'Clearly, then, C_{HL} [the computational system of the human language faculty] must include a . . . procedure that combines syntactic objects already formed. . . . The simplest such operation takes a pair of syntactic objects (SO_i, SO_j) and replaces them by a new combined syntactic object SO_{ij}. Call this operation *Merge*' (Chomsky, 1995a, p. 226). The elements that Merge operates on obviously include the lexical items that the computational procedure starts with and also the structures formed by the procedure itself. In this way a whole complex structure can be built up by Merge. The procedure is rather like building a model from a kit: you start with all the individual pieces and you glue them together; the part formed by gluing pieces one and two may be glued to that formed by gluing pieces three and four, and so on until the model is completed.

One might wonder why we start off with a specific set of lexical items rather than simply assuming that the computational procedure accesses the lexicon directly to fetch the words to build structures from. Like many other questions, this is an empirical issue that depends on seeing how 'correct' it is to assume an initial selection of lexical items. However, there are conceptual grounds which lend it support. Chomsky proposes the following analogy: 'Suppose automobiles lacked fuel storage, so that each one had to carry along a petroleum processing plant. That would add only bounded "complexity", but would be considered rather poor design' (Chomsky, 2000c, p. 12). If the computational process has access to the lexicon as a whole at any stage of the derivation, then it 'must carry along this huge beast, rather like cars that have to replenish fuel supply constantly' (Chomsky, 2000c, p. 13). The point is that the computational process is much

simplified if its operations are limited to a set of pre-selected lexical items. For example, it imposes a strict condition on the derivation in that, once all the selected elements have been included, no more can be added and the derivation can halt. If the whole lexicon were available to the computational procedure, then the derivation could potentially be applied over and over again and no decision about the grammaticality of the structure could be made.

Another argument can be advanced from an economy point of view. Economy of derivation suggests that the one that should be preferred out of all the possible derivations for a structure is the one with the fewest steps, other things being equal. Thus different derivations compete with each other, and the shortest one is preferred. But, if there are no limits on which derivations can compete, meaning that all derivations for any possible structure compete with each other, then the entire language will reduce to the single most economic derivation, perhaps a structure with just one word. This can be avoided by restricting the set of competing derivations to those which are based on an initial selection of lexical items: the derivation of one-word structures will not compete with derivations based on a set of five lexical items. To take a more realistic example, restricting competition to derivations based on an initial selection of lexical elements will prevent derivations for the sentence *men like beer* from competing with those for the sentence *those men like beer*. If these were to compete, a derivation for the first sentence would win as it contains fewer words and therefore fewer derivational steps.

Note that implicit in these arguments is the assumption alluded to above that all the selected lexical items must be used in each competing derivation; otherwise shorter derivations, which make use of just one of the selected items, would win. For this reason, the initial selection of lexical items cannot be viewed as a simple set, as we have to indicate how many times a given item is to be used in a derivation. For example, the following two sentences are derived by using the same set of words, but the word *the* is used a different number of times:

(11) a. The teacher hates the students.
b. The teacher hates students.

Obviously the derivations of these two structures do not compete with each other, otherwise only (11b) would be grammatical because its derivation is shorter. Because the frequency with which a given lexical item is to be used is important in distinguishing between these kinds of derivations, Chomsky refers to the initial selection of lexical items as a **Numeration**. Each Numeration will consist of the set of words to be used in a derivation and an indication of how many times each word is to be used. This is given by a numerical index assigned to each selected lexical item. Thus, the Numerations for the sentences in (11) would be something like:

(12) a. (The_2, teacher_1, hates_1, students_1).
b. (The_1, teacher_1, hates_1, students_1).

So far we have considered properties of the starting point of the computational procedure, the Numeration; what about the end point, the LF? As we have said,

each LF derived from a given Numeration must include all of the elements indicated in the Numeration the required number of times. It is at LF that each derivation is evaluated in terms of Full Interpretation: an LF which is not interpretable, or 'legible' as Chomsky terms it, was not produced by a valid derivation; only valid derivations yield grammatical structures. Chomsky describes a valid derivation as **convergent** and an invalid derivation as one which **crashes**: 'a derivation D *converges* if it yields a legitimate SD [Structural Description] and *crashes* if it does not' (Chomsky, 1995a, p. 171).

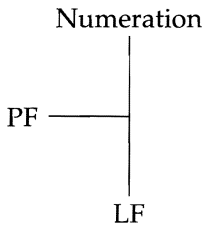
There will be more to say about the kinds of things that would cause a derivation to crash later on; however, one point can be made from our present position. At any point in the derivation, there can be any number of partially built structures which may be merged into a single structure at some later step. At LF, only a single structure can be interpreted as, for a set of disconnected partially built structures, it is not stated how such elements are to be related to each other for interpretation purposes: which is the complement of which; which, if any, are specifiers; etc.? Thus, although Merge can apply freely, if the derivation is not to crash by the time it gets to LF, all of the substructures will have to be merged into one. It is a minimalist assumption that the status of a derivation is determined only by the Bare Output Conditions imposed by the interpreting faculty of the mind, in this case the conceptual-understanding faculty. It therefore follows that no grammatical principles can apply during the derivation itself that serve to separate grammatical from ungrammatical derivations: a derivation is determined to be convergent or crashed only at LF. Note that it is not guaranteed that there will be any convergent derivation for all possible Numerations. A Numeration made up of 34 instances of the word *ago* will never produce a convergent derivation no matter how Merge is applied. This is not a problem as we wouldn't want the grammar to determine a grammatical structure based on these words. All the grammar should do is determine which derivations produce grammatical structures and which do not, and, if it does this accurately, then it is adequate.

Let us briefly summarize what has been said so far. The computational process starts with a Numeration of lexical items and builds these freely into structures using the operation Merge time and again. The final step in the derivation yields an LF which will be presented to the Output Conditions to be tested for its legibility. If it is a single valid structure containing nothing uninterpretable, the derivation has converged and a grammatical structure will have been derived. However, if the LF formed is not valid, but violates the Output Conditions, the derivation will crash and the derived structure is deemed ungrammatical.

Let us now turn our attention to PF. The derivation of PF cannot be totally divorced from that of LF as the two interface levels must be connected in the appropriate ways across the bridge between sounds and meaning: something that means *John loves Mary* is not going to be pronounced as *the cat sat on the mat*, for instance. Moreover the phonological features which are interpreted at PF are introduced by the lexical items which form the Numeration, and certain aspects of grammatical structure do affect the pronunciation. The derivations of LF and PF share at least some part of the computational process. However, a PF representation is a very different object to an LF representation and is constructed largely by different mechanisms which manipulate phonological rather than grammatical or

semantic material. At a certain point, then, the derivation must split, with all the phonetically relevant material being sent in one direction and all the remaining material being sent in the other. From this point on, the two parts of the derivation proceed independently towards their final destinations: a fully formed LF and PF. This can be represented in the following way:

(13)



Note that the line from the Numeration to LF is straight while that to PF shoots off in a different direction. This represents the fact that the computations which form PF objects are of a different nature from those that build syntactic structures. LFs, on the other hand, are syntactic structures built in a uniform way.

The point at which the derivation splits is known as **Spell Out** – it is where the derivation is 'spelled out' in terms of its outward physical form. Given that the external representation of the derived structure is determined by this point, the computation which continues after Spell Out en route to LF is covert: its operations will not affect the shape of PF. Thus, as in GB, there is an overt part of syntax that precedes Spell Out, and a covert part of syntax that follows Spell Out.

The principle of Full Interpretation also applies at PF, relativized to the phonetic material that PFs are constructed of, and as such a derivation which presents the sensorimotor systems with material which is not phonetically legible will crash. A whole derivation therefore only converges if it converges both at LF and at PF.

Given that Spell Out marks the point at which the derivation branches off to PF and LF, and therefore that any step prior to its application will be visible and those following leading to LF will be invisible, one might be forgiven for thinking that Spell Out is equivalent to S-structure, which occupied a very similar position in GB Theory. However, this is not the right way to view things. In GB, the levels of representation, D-structure, S-structure, LF and PF, were all fully formed objects judged for well-formedness by various criteria, depending on which modules of the grammar were applicable to them. So D-structure conformed to θ-Theory and S-structure conformed to Case Theory, etc. The individual steps in the minimalist derivation, including the one at which Spell Out happens, do not yield fully formed structures but only ones that are partially built. Furthermore, they are not subject to evaluation for well-formedness. It is only at the very end of the derivational processes that the constructed objects (LF and PF) are evaluated in terms of their legibility. Thus, while S-structure was a level of representation in GB, Spell Out is merely a step in the derivation, and so the two are entirely different concepts.

As the well-formedness of the structures under construction is not evaluated until after they are completely built, it would be reasonable to enquire how we

know when to apply Spell Out. The strict answer is that there are no restrictions on where Spell Out may apply. However, only if it is applied at the right point will the derivation converge. For reasons that will become clearer later, under minimalist assumptions Spell Out must apply as soon as possible. However, two things must be achieved before Spell Out can apply. The first is that all overt elements in the Numeration must have been incorporated into the structure. This follows straightforwardly from the fact that lexical elements have both semantically relevant and phonetically relevant features; it is Spell Out which divides these and sends them to their appropriate destinations. If a lexical element is inserted after Spell Out, on the track to either LF or PF, material relevant for one interface will arrive at the other, where it will be illegible and thereby cause a crash. The second precondition before Spell Out can apply is that a single structure must be built. This follows from the fact that the path to PF does not involve syntactic mechanisms which are responsible for structure building (i.e. Merge) and therefore, under the assumption that PF is no more able to interpret a set of unconnected structures than is LF, all substructures must be combined into a single structure before it sets out on the track to PF.

To summarize, the basic computational process takes pre-selected lexical items and builds them into a structure by successive applications of Merge. At the point of Spell Out, the derivation splits and phonetically relevant features are sent one way and grammatical and semantically relevant features another. The computational procedure continues, applying similar processes to build a fully constructed LF and phonetically relevant processes to build the PF. At these points, the resulting objects are presented for evaluation for their interpretability: if the LF is semantically interpretable and PF is phonetically interpretable, the derivation converges; if not, it crashes.

BASIC CONCEPTS OF THE MINIMALIST PROGRAM

Numeration: A selection of lexical items involved in building a given sentence plus an indication of how many times they are to be included in a structure, which constitutes the starting point of the structure-building process.

Computation: The step-by-step procedure which builds the elements of the Numeration into a fully formed LF and PF. It consists of **Merge**: an operation which reiteratively makes new structures out of a combination of two elements from the Numeration or from the structures already formed.

Spell Out: The point at which the derivations split, sending off relevant material to LF and PF.

Convergence: A derivation is evaluated only at its end points, LF and PF. If a legitimate structure has been built, i.e. one which satisfies the Bare Output Conditions (Full Interpretation), then the derivation converges; otherwise it crashes. Convergent derivations produce grammatical structures, crashes do not.

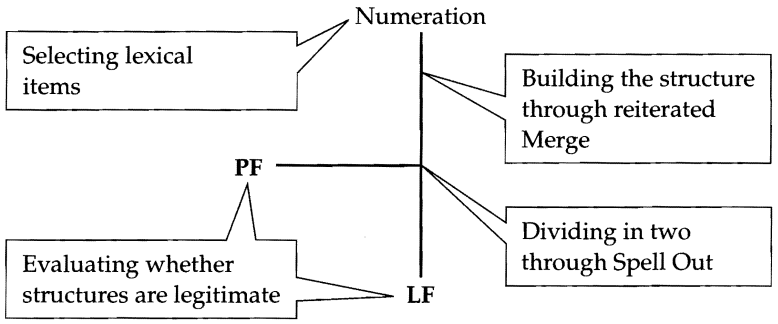


Figure 7.3 Basic structure of the Minimalist Program

EXERCISE 7.1

Given the words *that, man, loves* and *cooking*, how many binary combinations (mergers) of these are possible and how many would converge?

7.3 Phrase structure in the Minimalist Program

The previous section briefly outlined the working of the computational system that provides the central component linking LF and PF. This section fills in some of the gaps concerning how appropriate phrase structures can be constructed by Merge without the assumption of extra grammatical mechanisms to steer this process.

Chapter 3 introduced X-bar Theory as a framework for accounting for phrase structure phenomena. There is much empirical evidence to support the structures that X-bar Theory produces and, within a GB framework, there were many conceptual advantages for adopting it. However, within the MP there is, strictly speaking, no room for X-bar principles. X-bar structures are clearly not imposed on grammatical objects at the interface levels; there is no interpretative semantic or phonetic reason why language structures should look like those produced by X-bar principles. If there is an X-bar component in the grammar, it would therefore have to apply during the computational procedure. Yet the basic assumption of the MP is that no principles apply other than at LF and PF, and so every attempt should be made to eject X-bar Theory from the grammar. In order to avoid throwing the baby away with the bathwater, however, we might hope to be able to produce X-bar-type structures in some way, thus holding on to their empirical advantages, without having X-bar Theory as a module of the grammar.

There are other reasons for abandoning traditional X-bar Theory. One concerns the binary branching of structures. From an X-bar position, this had to be imposed as an extra stipulation through the Binary Branching Condition (Kayne, 1984). However, the condition can be easily 'built in' to Merge if this is taken to be an operation which combines two elements into one. As a result, all structures built by Merge will be necessarily binary branching, no more and no less. But this is different to the Binary Branching Condition of X-bar Theory, which states

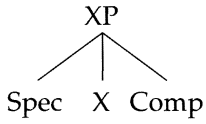
that structures can be *at most* binary branching, i.e. structures are allowed to be unary branching as well. For example under the assumption that pronouns are 'intransitive' determiners (in that they take no NP complement), they might be given the following structure:

(14)

But such a structure makes no sense from the perspective of Merge: what was merged with the pronoun to form the D′ or the DP? Thus such structures as (14) could only exist if we allow some other mechanism in the grammar to build them. The minimal assumption is then that such unary structures do not exist. This problem clearly highlights an incompatibility between X-bar Theory and basic minimalist assumptions.

There are other problems that are internal to X-bar Theory itself. For example, as various alternative proposals for X-bar Theory during the 1970s testify (e.g. Chomsky, 1970; Siegel, 1974; Jackendoff, 1977), the theory itself does not restrict how many projection levels there can be, though it has become the norm since the 1980s to assume that the second projection level is maximal; in other words, XP = X″. However, this is merely a stipulation based on the observation that this is the smallest number of projection levels that can capture most phrase structure phenomena observable in the world's languages. There is no reason that necessarily follows from X-bar Theory itself why this should be. Clearly this is a shortcoming of the theory.

X-bar Theory faced another problem as a result of the developments of the later 1980s. The intermediate projection X′ obviously forms a necessary part of X-bar Theory, as it is X′ that distinguishes the structural position of specifiers and complements. Without the X′, specifiers and complements would both be sisters to the head and hence structurally indistinguishable:

(15)


However, with developments such as the DP Hypothesis a lot of the empirical evidence for the existence of X′ was gradually undermined. For example, what had been taken to be cases of N′ coordination and pronominalization, as in (16) below, were reanalysed as cases of NP coordination and pronominalization within a DP structure, as in (17):

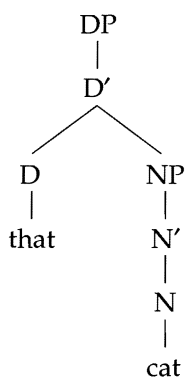
(16) a. [$_{NP}$ the [$_{N'}$ man in white] and [$_{N'}$ woman with a hat]]
b. I read this book about linguistics, not [$_{NP}$ that [$_{N'}$ one]].

(17) a. [$_{DP}$ the [$_{NP}$ man in white] and [$_{NP}$ woman with a hat]]
b. I read this book about linguistics, not [$_{DP}$ that [$_{NP}$ one]].

In the phrases *the man in white* and *this book about linguistics* the sub-strings *man in white* and *book about linguistics* are constituents as they can be coordinated with another like-constituent and can be replaced by a pronoun. Before the DP Hypothesis, this constituent could only be an N′, as it was smaller than the NP but bigger than the noun. But once the DP analysis was introduced, this constituent could be taken to be a full NP and thus the data no longer support the existence of a constituent smaller than a phrase. Moreover, while we find many instances of phrasal and head movements (e.g. passivization and auxiliary inversion), there are no instances of X′ movement. Such observations led Chomsky to claim that the single-bar X′ is grammatically inert, playing no role in grammatical processes such as movement, coordination or pronominalization: 'bare output conditions make the concepts "minimal and maximal projections" available to C_{HL} [the human linguistic computational system]. But C_{HL} should be able to access no other projections' (Chomsky, 1995a, p. 242). The problem then is to account for the existence of a grammatical element, which is clearly necessary for accepted structural description, but which itself plays no role in any grammatical process.

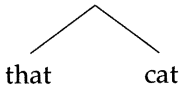
In 'Bare phrase structure' (1995b), Chomsky addressed these issues by proposing a system which achieves the relevant X-bar-like structures without the assumption of an extra X-bar module in the grammar. His starting point was to assume that a minimalist structure-building process would make use of the resources provided by the Numeration, and nothing else. This property Chomsky calls 'inclusiveness': 'any structure formed by the computation is constituted of elements already present in the lexical items selected for N[umeration]; no new objects are added' (Chomsky, 1995b, p. 393). Clearly standard X-bar Theory is not inclusive as it adds material to the lexical resources that the structure is built on. For example, consider the following structure:

(18)



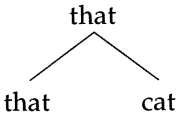
For this structure, the Numeration would contain the words *that* and *cat*. But X-bar Theory makes us erect N, N′, NP, etc. over these items, and these labels are clearly not part of the Numeration. An inclusive system would just take the two words and join them together as a unit:

(19) that ⟶ cat



However, given that the unit formed by this process has properties of its own, i.e. it is a DP, (19) is not adequate as a description of the structure. What is lacking is the label. But, if an inclusive system only draws on the elements of the Numeration, one of these must be utilized as the label. This step gives the notion of **head** a proper treatment within this system: the head is the element which is selected to be the label of the phrase formed by merging two elements. Thus, a better representation of the result of the merger given in (19) would be:

(20)



This structure looks very different from the standard X-bar structure such as the one in (18), but in fact it isn't. Clearly the determiner is indicated as the head of the phrase, as it projects and the noun does not. Therefore this structure represents a DP, a phrase headed by a determiner, just as the one in (18) does. Furthermore, the noun is the sister to the head and hence it occupies the complement position on the standard definition that the complement is the sister of the head.

We can go even further than this. Chomsky claims that bare phrase structure represents nearly everything represented by an X-bar structure, assuming the following definitions:

- A head (zero-level projection) is something which is not a projection.
- A phrase (maximal projection) is something which does not project any further.

The point is that the notions *head* and *phrase* can be given a contextual definition: defined in terms of the structural context they occur in. With this in mind, consider the word *cat* in the bare phrase structure (20). As this is not projected from anything, it is a head and, as it projects no further, it is also a phrase at the same time. This is exactly what is represented in the X-bar structure in (18) by means of building N, N′ and NP on top of the word *cat*. Now consider the word *that*. This seemingly appears twice in (20), but this is an illusion due to the limitations of the orthographic representation being used. As we know, lexical items are a complex collection of features, some semantic, some phonological and some grammatical. While the semantic and phonological features are important for the word at the bottom of the structure, they are not important for the element used for the label of the structure. What is important for the label is the categorial features. The lower instance of *that* is the head: it is not projected from anything. The higher instance is a phrasal projection: it is projected, but it projects no further. Once again, then, this represents exactly what is represented by the tree in (18).

Figure 7.4 shows how the two structures relate to each other.

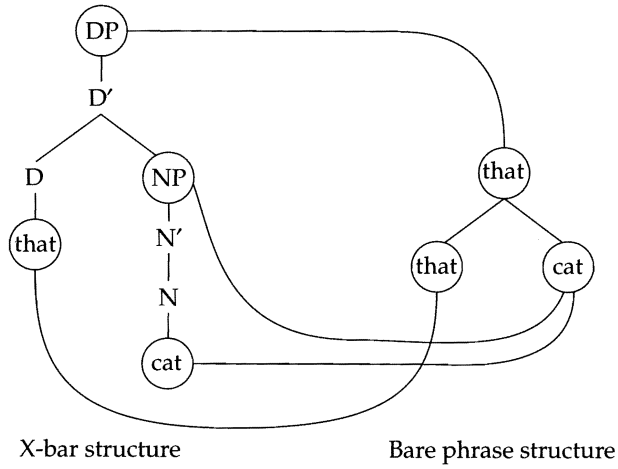
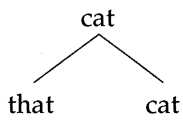


Figure 7.4 Relationship between X-bar structures and bare phrase structures

In other words, the bare phrase structure states that the word *cat* is the head of an NP which sits in the complement position of a determiner *that*. This determiner is itself the head of a DP. The only difference between what is represented in the two structures concerns the X′ element: the bare phrase structure does not claim that either *cat* or *that* has X′ status. Yet, as we have seen, X′ is grammatically inert and hence there can be no empirical evidence to show whether something is an X′ or not, and so this difference reduces to a formality with no empirical consequences.

But why choose the determiner as the label rather than the noun? Why should the structure not be:

(21)



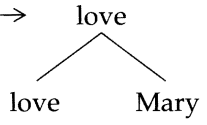
The question is empirical: which of the two elements is the head? To decide this means knowing the lexical properties of the elements involved: determiners take NP complements but nouns do not take determiner complements (or specifiers). However, the correct selection of the head is not a restriction we have to place directly on Merge, allowing only the formation of (20). Suppose that in a derivation the noun is selected to be the label, as in (21). Then the derivation will crash because the elements are not placed in the relevant structural relationship to be interpretable at LF: the determiner requires an NP complement, but in (21) the determiner is sitting in a (pre-head) complement position (sister to the head) of a phrase headed by a noun. Only if the determiner is selected to be the label, as in (20), can the derivation converge, and hence it is correctly predicted that (20) is grammatical while (21) is not.

So far we have dealt with heads and phrases, but what about intermediate projections? The contextual definition of bar projections allows one further case, that of something which is neither a head, because it is projected, nor a phrase,

because it projects. This is exactly the case of an intermediate projection: an X′ (intermediate projection) is neither a head nor a phrase.

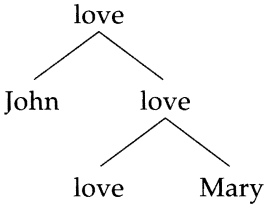
To see how such things arise, let us consider the first few steps of the derivation of a sentence. Suppose the Numeration contains, amongst other things, the lexical items *love*, *John* and *Mary*. An initial step might be to take the verb and merge it with one of the nouns. As verbs take nominal complements and nouns do not take verbal complements, it follows that only if the verb is projected will the derivation converge. Thus the first step in the derivation might be:

(22) love Mary ⟶



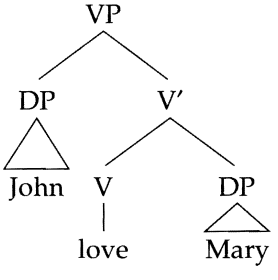
At this stage in the derivation the projection of the verb is a phrase, as the verb projects no further than this. The next step is to merge the other argument of the verb with the VP constructed so far. Supposing we once again select the verb as the label, the resulting structure is:

(23)



This places *John* in the specifier position: sister to the intermediate projection, daughter of the maximal projection. As verbs take their external arguments in their specifier position, the structure is able to be interpreted and hence converge. At this point in the derivation the label at the top of the structure is the maximal projection of the verb, i.e. it is the VP. However, the first projection of the verb is no longer maximal and hence it is not a phrasal projection. Clearly it is not a zero-level category either, as it is projected. From this perspective, then, intermediate projections are seen to be nothing more than projected elements which were maximal projections at one point in a derivation, but which are projected at a later point: 'From a representational point of view, there is something odd about a category that is present but invisible; but from a derivational perspective, . . . the result is quite natural, these objects being "fossils" that were maximal (hence visible) at an earlier stage of derivation' (Chomsky, 1995b, p. 435). The structure in (23), therefore, is the exact equivalent of the more familiar:

(24)



EXERCISE 7.2

Using Bare Phrase Structure notation, what would the following phrases look like?

about it
news about it
some news about it
hear some news about it

Do you foresee any problems for the analysis of the following?

some awful news about it

A brief recap of the system just outlined will show how it addresses the problems discussed at the beginning of this section. The operation Merge is extended slightly to include not only the joining together of two elements, but also the selection of one of these as the label of the newly formed structure. This selection is free, but it has consequences that affect the interpretability of the final structure. Selecting one of the elements to project will immediately make the other a maximal projection sitting in either the complement or specifier of the projected element. This straightforwardly accounts for a common stipulation in X-bar Theory, that the only non-phrasal positions in a phrase are the head and its non-final projections. Thus, all complements and specifiers are maximal projections, something which follows directly from a contextual definition of projection levels. The lexical properties of the merged elements may or may not be satisfied under this arrangement, but only if they are will the final structure converge.

As we have mentioned, the very process of selecting a label builds in to the system the notions of head and projection that were central to X-bar Theory. The projection level of any element in the structure can be determined contextually by looking to see if it is projected or if it projects further. Thus, all of the core notions of X-bar Theory are incorporated into the system. As Merge is a binary operation joining together two elements, it produces only binary branching structures without having to stipulate this as an independent condition on structures. Under the contextual determination of projection levels, however, the need for unary branching structures is completely done away with, hence addressing this problematic issue. A single element might be a head and a maximal projection at one and the same time without the need to stipulate different projection levels to this effect.

Also note that the contextual definition of projection levels means that we do not have to stipulate how many projection levels there are, as X-bar Theory does. From the Bare Phrase Structure perspective there can only be three levels: the word, which is not projected; the phrase, which does not project; and the intermediate projection, which both projects and is projected. There can be no other kind of element contextually defined in these terms, and every element in a structure is defined as at least one of these things.

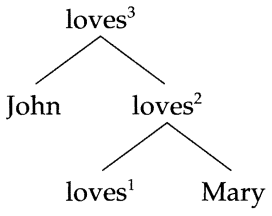
Finally, consider the status of the intermediate projection X′. This element is present in structures, but inert. This follows from the fact that when such elements

emerge in a structure they are defined as neither one thing nor another: they are not heads and they are not maximal projections. Supposedly, it is this property that renders them invisible to the system, and hence they can play no further role in the derivation of a structure.

THE BARE PHRASE STRUCTURE SYSTEM

Merge (extended version): An operation which takes two elements and joins them in a single structure, selecting one as the label of the new structure.
Projection levels are represented contextually under the following definitions:

- An element which is not projected is a zero-level category.
- An element which does not project is a maximal projection.
- An element which is neither a zero-level category nor a maximal projection is an intermediate projection (X′), e.g.:



Here both *John* and *Mary* are zero-level categories (because they are not projected) **and** maximal projections (because they do not project). However, *loves*^{1} is a zero-level category and *loves*^{3} is a maximal projection. *loves*^{2} is an intermediate projection, being neither a zero-level category (as it is projected) nor a maximal projection (as it projects).

7.4 Thematic roles and structural positions

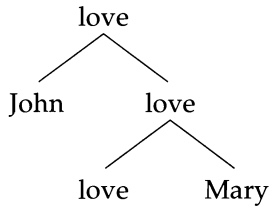
In GB it was assumed that predicates are related to their arguments via a process of θ-role assignment. Internal θ-roles were assigned to complements and the external θ-role was assigned to the subject. Theta Theory, the module which governed the process of θ-role assignment, was assumed to apply at D-structure. However, in the MP there is no such thing as D-structure and indeed, a process of θ-role assignment that operated in the derivation of a structure would be something extra, not something imposed by the Bare Output Conditions. This raises the question of how the effects of Theta Theory are to be handled under minimalist assumptions.

If there is no process of θ-role assignment, then the interpretation of arguments with respect to their predicates must take place at the LF interface. One possibility

is that arguments are interpreted in a particular way due to the structural positions they occupy: an argument in the specifier position of a certain type of verb might be interpreted as agent, for example, or one in complement position of another type of verb might be viewed as the goal. This idea stems from work by Hale and Keyser (1993), drawing heavily on Baker's (1988) **Uniform Theta-Role Assignment Hypothesis**, which states that θ-roles are assigned uniformly in all languages. That is, there are certain positions to which certain θ-roles are assigned to universally. Once certain universal positions are associated with certain θ-roles, however, we can give up the notion of θ-role assignment as a grammatical process and view the phenomenon as the interpretative consequence of an argument occupying a given position.

One question that arises concerns how structural positions are to be defined under the Bare Phrase Structure assumptions outlined above. Chomsky points out that under these assumptions notions of complement and specifier can be easily defined in more or less the traditional way. The complement is the sister to the head and the specifier is the sister to the X′, daughter of the maximal projection of the head. Consider the bare phrase structure given above in (23) once more:

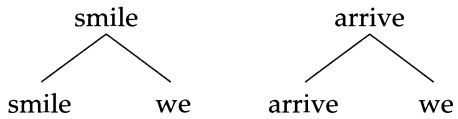
(25)



In this VP, *Mary* is the complement as it is sister to *love*, the projecting head. *John* is the specifier, daughter of the last and therefore maximal projection of the projecting head. The complement is therefore the result of the first merger with a head, and the specifier is the result of a subsequent merger. Given the interpretation of this configuration, with *John* as experiencer and *Mary* as theme, we might therefore conclude that an argument in complement position is interpreted as theme or patient and an argument in specifier position is experiencer or agent.

One problem for this assumption, however, is to distinguish between intransitive verbs such as *smile* and unaccusative verbs such as *arrive* (see chapter 4, section 4.1.1); while both of them have only one argument, intransitives have an agent argument and unaccusatives have a theme argument. In GB the assumption was that intransitives placed their argument in subject position and unaccusatives placed their argument in complement position, though it might later move to subject position for reasons of Case. But, according to what we have said, the first merger with a head places an argument in complement position, and thus it is hard to understand how single argument verbs could differ in this way:

(26)

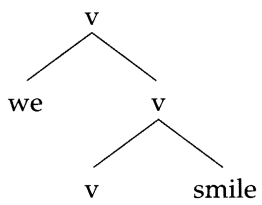


One solution might be to assume that intransitive verbs have some other element merged with the verb first so that the agent argument is merged into specifier position. This element might be something akin to a cognate object; possibly the verb itself might be abstract, taking its form from its object via a movement. Another possible solution is suggested by observations of causative verbs which take a VP complement and have an agentive subject (see chapter 4, section 4.1.3):

(27) John made [$_{VP}$ Bill clean the windows]

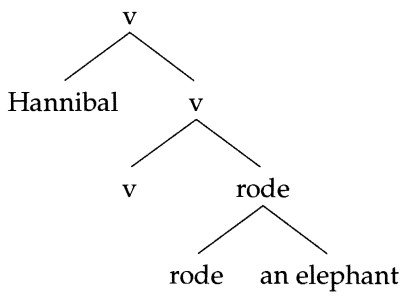
If the Uniform Theta-Role Assignment Hypothesis is taken strictly, it might be argued that an argument is interpreted as agent when it is in the specifier position of a verb with a VP complement. The agent of an intransitive verb must therefore be the subject of an abstract verb, taking the intransitive verb as its complement, rather like a VP shell structure:

(28)



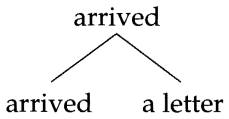
This abstract verb is often called a **light verb** and is represented as in (28) with a lower-case 'v', reserving the capital 'V' for the lexical item. Under Bare Phrase Structure assumptions, the verb *smile* is a phrase in the complement position of the light verb. As the light verb has a complement, the argument *we* will be merged into its specifier position and hence be interpreted as agent. If this analysis generalizes to all agents, a typical transitive verb will first merge with its object and then the VP formed in this way will merge with a light verb; finally the agent is merged with this structure:

(29)



Verbs lacking an agent, such as unaccusatives, may lack a light verb and hence there will be no agent position for these:

(30)



THETA THEORY IN THE MINIMALIST PROGRAM

The Uniform Theta-role Assignment Hypothesis (Baker, 1988):

θ-roles are assigned to uniform positions in all languages.

Assuming this allows us to associate certain positions with certain θ-roles, e.g.:

- agent = specifier of v taking a VP complement
- theme = complement of V

Arguments are then interpreted at LF as bearing the θ-roles associated with the positions they are merged into.

EXERCISE 7.3

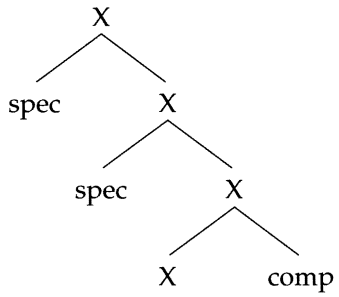
If we accept that agents are merged into the specifier of a light verb with a VP complement and that experiencers differ in that they are merged into the specifier of the thematic verb they are related to, what will the structure of the following sentences be, assuming that in the first the subject is an agent and the second an experiencer?

John broke the window.
John broke his arm.

7.5 Adjunction

So far we have discussed the complement position, the result of the first merger with a head, and the specifier position, resulting from a subsequent merger. It remains to discuss adjunction, introduced in chapter 3. This, however, turns out to be highly problematic from a Bare Phrase Structure point of view. One might think that adjunction is simply the result of mergers which take place between the first and last operation of Merge, placing all adjuncts as sisters to intermediate projections, as the X-bar Theory of the early 1980s assumed. However, this would mean that we could not have an adjunct unless there were a complement and a specifier, otherwise the intended adjoined element would be placed in complement or specifier positions. Besides, for technical reasons concerning movement, Chomsky claims that all mergers after the first place elements in specifier positions. This means that, contrary to standard X-bar assumptions, structures can have multiple specifiers:

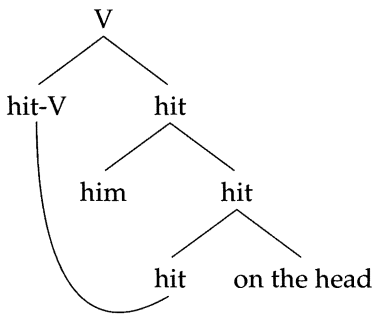
(31)



According to the Bare Phrase Structure analysis, the second spec element in (31) is not in an adjunct position: when it was merged into the structure, note, it was merged into the specifier position, under the maximal projection of the head.

But under these assumptions, then, how are we to treat adjuncts? There are several possibilities. One is to assume, along with Larson (1988), that there are no adjunction positions, but that adjuncts are merged into a structure early and hence are 'inner complements' of a head. This necessitates a VP shell structure (see chapter 3, section 3.6.4), with the verb being first merged with the adjunct and then moving to higher V-positions where the arguments will be merged:

(32)



One piece of evidence in favour of this assumption is that an object may bind an anaphor inside an adjunct, indicating that the object c-commands the adjunct:

(33) I gave the boys presents on each other's birthdays.

However, under these assumptions it is impossible to maintain a strict version of the Uniform Theta-Role Assignment Hypothesis, as arguments would be merged into different positions depending on how many adjuncts there are: if there are no adjuncts, an object would be merged into complement position and, if there is an adjunct, the object would be merged into the specifier position. The more adjuncts there were, the higher the specifier position that the object would have to be merged into.

More recently, Chomsky has voiced a dissatisfaction with previous treatments of adjunction: 'There has never, to my knowledge, been a really satisfactory theory of adjunction' (Chomsky, 2001b, p. 15). He notes that for most purposes other than semantic interpretations adjuncts behave as though they are not there: they don't alter the syntactic nature of the element they are adjoined to by projecting

their own properties as do heads, nor do they fulfil any selectional requirement of the head as arguments do. He also points to evidence that adjuncts behave very differently from arguments in other ways. For example, consider the following sentence:

(34) Which picture of John that Bill liked did he buy?

The relevant observation is that *John* cannot be the antecedent of the pronoun *he*, but *Bill* can. The standard explanation of why *John* cannot be the antecedent is that in such cases, which involve movement, the referential relationships are judged from the original position of the moved element. Given that the interrogative phrase originated in the object position, the subject pronoun *he* would bind the r-expression *John* if they were co-referential, in violation of Condition C. However, if this explanation holds for *John*, the complement of the noun, why does it not also hold of *Bill*, which is in the relative clause adjoined to the noun? It seems that the adjunct is not evaluated from the lower position, as if it were never in this position. This has led some to suggest that adjuncts are added to structures late in the derivation, after the movement has taken place in (34) for example. For technical reasons, this is not a solution that Chomsky wants to adopt, however. Although it is far from a full theory of adjunction, he suggests that adjuncts are included not by Merge but by an operation which replaces the element that is adjoined to with a pair of elements constituted of the replaced element and the adjunct. This pair behaves exactly like the element that is replaced and all existing relationships involving this element are unaltered. For syntactic purposes, therefore, it is as though the adjunct is not there. Clearly, however, for interpretative purposes, at LF and PF, the adjunct must be visible. Therefore Chomsky assumes two interpretative procedures, one which applies to LF and one which applies to PF. The LF interpretative procedure replaces the pair <adj, XP> with structures of the same kind that complements and specifiers are involved in, so that adjuncts can be interpreted along the same lines, entering into semantic relationships involving c-command, for instance. The PF procedure is an instruction concerning where in the linear string the adjunct is to be pronounced. For this, Chomsky assumes that an adjunct is pronounced wherever the phrase that it is adjoined to is pronounced.

It is difficult to know how to represent this conception of adjunction in terms of a phrase structure tree, and Chomsky does not develop any conventions in his brief discussion of these ideas. However, tree diagrams are merely convenient ways to represent theoretical ideas and have little theoretical import of their own; there are many other ways one can think of to represent the results of a set of mergers. One way that Chomsky has favoured in recent years is in terms of sets. A merger of two elements produces a set made up of the set of the two merged elements and the label, chosen from these:

(35) the cat → {the, {the, cat}}

The object formed by this can clearly be distinguished from the result of adjunction, which as we said produces a simple pair:

(36) hardly move → <hardly, move>

This pair has no label, but simply behaves as though the adjunct is not present, and hence for syntactic purposes it has all the properties of the verb. No doubt, one could render this distinction in terms of a tree diagram. However, this is not something we will venture to develop here.

Clearly there is much more to be said about the syntax of adjunction than Chomsky has so far ventured – why, for example, are there no adjuncts between predicates and their internal complements? Such questions must await a more fully developed theory.

ADJUNCTION

The VP shell approach: Adjuncts are merged into 'inner complement' positions, hence there is no such thing as adjunction. Under this analysis, arguments will be merged into higher specifier positions and the verb will raise to higher verbal positions.

The no-merger treatment: Adjuncts are included in a derivation by an operation which replaces an element X with a pair <adj, X>. This pair has all the properties of X and hence it is as though the adjunct is not there. At LF and PF there are special principles which tell us how the adjunct is to be interpreted, i.e. forming a constituent with X at LF and being pronounced where X is pronounced at PF.

7.6 Linear order

Due to the layout of pages, the structures drawn so far seem to reflect the linear ordering of elements in a structure; indeed, tree diagrams are traditionally assumed to represent both hierarchical relationships, in terms of **dominance**, and linear relationships, in terms of **precedence**. However, Chomsky has been moving away from this idea since around 1996, claiming that linear order is a property of PF while hierarchical structure is the sole property of LF. In other words, linear order is forced onto language by the fact that it must be externalized in ways limited by the human linguistic apparatus: the vocal tract for speech or hands etc. for sign language. Indeed, speech is far more linear than sign language, in which certain linguistic elements can be expressed simultaneously by a gesture of the hands being placed in a certain part of the signing area, for example. One might conjecture that if humans were telepathic, there would be no need for linearization at all.

We might conceptualize an LF, then, as more like the kind of mobile that parents hang above a child's crib, the constant movement of which will occupy the child long enough for them to have a cup of tea. For such an object linear order is meaningless. At any point two elements connected to the same node might be to the left and right of each other, or in front of and behind each other. However, hierarchical relationships remain constant no matter what positions the individual elements occupy in a three-dimensional space. It is these relationships, and others

defined in terms of them, such as c-command, which are important for semantic interpretation. Hence, Chomsky claims that semantically motivated processes should not make reference to linear relationships such as 'left' or 'right'.

Of course, hierarchical arrangement does have some effect on possible linear order. If one takes a child's mobile and places it on a flat surface, the individual objects will fall in a linear arrangement, but not all linear orders will be possible. So, while hierarchical structure may limit possible linear orders, it does not determine them. What determines linear order, within the limits set, is phonologically motivated and hence not part of syntax. This means that something like the head parameter, however it is to be construed in these terms, is not a syntactic parameter. This is well as this is one parameter that does not seem to be derived from properties of the functional elements of a language, which is where Chomsky assumes syntactic differences to be situated.

There is another approach to word order which Chomsky often refers to, based on work by Kayne (1994), which assumes that word order reflects hierarchical structure in a direct way. From this point of view, if one element c-commands another, then the first precedes the other. From this perspective, there is just one underlying word order (assuming that arguments are uniformly included into structures) and hence no head parameter at all. Different word orders must be the result of differences in movement possibilities in each language. But because of these added complexities Chomsky seems disinclined to accept this approach; 'An alternative . . . is that order reflects hierarchy. That approach eliminates the head parameter, but at the cost of introducing many others (options for movement required to yield the proper hierarchies) and also some technical complications' (Chomsky, 2001b, p. 7). To some extent, Chomsky's preferred approach of seeing word order as part of the derivation from Spell Out to PF redefines the subject-matter of syntax. Traditionally, word order has been seen as a central part of syntax, but under these assumptions it is somewhat marginalized. However, linguistic phenomena, like any natural phenomena, do not come ready sliced into neat compartments. What appears is a rather chaotic jumble of things, only some of which are properly considered linguistic. It is up to the theory to impose compartmentalization on this chaos; the extent to which this helps us to understand the world is a mark of how good or bad the theory is. Just because tradition has happened to compartmentalize word order phenomena as part of syntax is therefore no reason why we should continue to do so.

LINEAR ORDER

The Universal Word Order approach (Kayne, 1994): Word order reflects structure. If an element c-commands another then it precedes it. All languages therefore have the same underlying word order, and differences are brought about by different movements taking place.

The PF approach: Word order is a PF phenomenon and is not part of the syntactic component. Thus it is something that is imposed on a structure on the path from Spell Out to PF. LF concerns hierarchical relationships only and notions of left and right are undefined.

This chapter has outlined some of the basic aspects of Chomsky's current approach to issues in syntactic structure. These were largely worked out in Chomsky's 'Bare phrase structure' (1995b) paper and have remained fairly stable since. As we have seen, the Bare Phrase Structure approach has a number of points in its favour in comparison with standard X-bar Theory. However, questions obviously remain and the work is far from completion. The relationship between word order and the rest of the syntactic system is one area that remains to be investigated more fully, and the theory of adjunction has only really been put on the agenda in Chomsky's most recent work. The next chapter turns to movement phenomena, where there are even more radical departures from standard assumptions.

Discussion topics

1 To what extent would you consider the following aspects of the GB model to be part of an 'optimal design' for human language?

   the Case Filter
   X-bar Theory
   movement

2 Using Bare Phrase Structure terminology, define the following syntactic concepts: head, projection, phrase, complement, specifier, adjunct.

3 Could Bare Phrase Structure be made any more minimal? Which of the following would you consider to be indispensable: labels; the notion *head*; structure; the assumption that Merge is a binary operation?

4 List the advantages and the disadvantages of the Bare Phrase Structure system over standard X-bar Theory. Is the rejection of X-bar principles by the MP a positive step?

5 If linearization is excluded from syntactic phenomena by definition, to what extent has the MP defined away its own subject-matter?

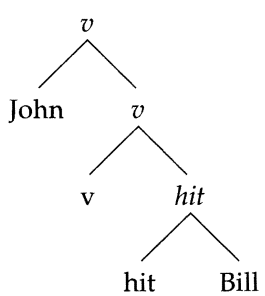
8 Movement in the Minimalist Program

The previous chapter introduced the minimalist approach taken by Chomsky towards issues in basic structure, mentioning movements that are also assumed, but not going into details. This chapter considers these movements more thoroughly, placing them in the general framework, discussing their nature and demonstrating how they work.

8.1 Functional heads and projections

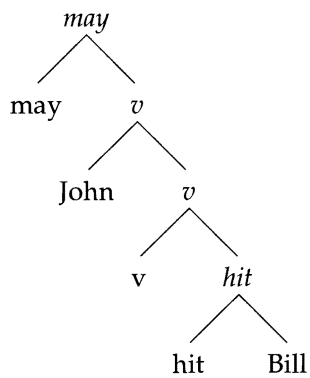
Chapter 7 deliberately steered clear of the functional structure usually assumed to be erected over the VP, namely the IP and the CP. How can these be incorporated into a Bare Phrase Structure analysis? Let us start with the inflectional elements. Lexical inflections, such as modal auxiliaries, will be included along with the verb and its arguments in the Numeration. Suppose the point in a derivation has been reached where the verb and its arguments have been merged into a single VP, which may or may not involve a light verb depending on properties of the main verb: if there is an agent, then there will be a light verb. As trees become more complex the fact that lexical elements and their projections are orthographically the same can be confusing, and therefore we will introduce the convention that projections will be represented in italics:

(1)



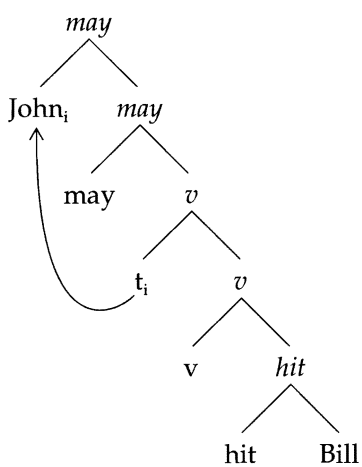
We now want to merge in the modal from the Numeration, say *may*. As inflectional elements take VP complements, this means straightforwardly merging the modal with the structure built so far and labelling it with the modal:

(2)



So far, so good. But in English the process does not stop here. In particular the subject moves out of the VP into the specifier of IP, as seen in chapter 3, section 3.4. Consider the structure that results from this movement:

(3)



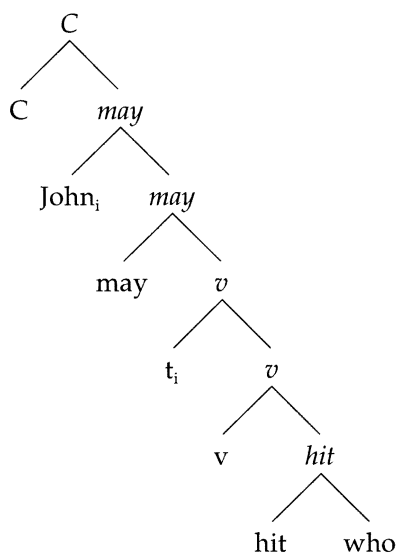
In many ways this movement is very similar to the Merge operation that we have been discussing. We take two elements, *John* and the structure built so far, and put them together to form a structure containing both. One is chosen to be the head, in this case the modal, and so the moved element is in the specifier of the IP. The only difference between this movement and a simple merger is that with merger the two elements that are merged are completely independent of one another, whereas here one of the two elements originates within the other. Chomsky introduces the terms **External Merge** to talk of the merger of two elements that are external to (separate from) each other, and **Internal**

Merge to talk of the merger of two elements, one being internal to the other. 'Merge of α, β is unconstrained, therefore either *external* or *internal*. Under external Merge α and β are separate objects; under internal Merge, one is part of the other, and Merge yields the property of "displacement"' (Chomsky, 2001b, p. 8).

External Merge results in two independent parts of the structure being joined together in a single structure; Internal Merge results in phenomena that we have been calling movement. Movement is then seen as a combination of one element with another element from a different part of the tree.

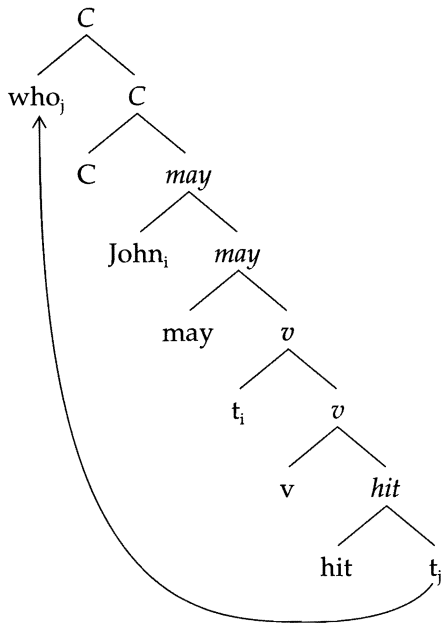
Note that both kinds of merger play a role in the building of the structure and so movement does not apply to a static structure which has already been built, as in Government/Binding (GB). Since the Minimalist Program (MP) has no D-structure and S-structure levels between which movements take place, each movement is a single step in the process moving from the Numeration towards Logical Form (LF). This can be made clear by considering a further movement. Imagine a partially built structure similar to (3), but having *who* instead of *Bill* in object position. Suppose there is an interrogative complementizer in the Numeration, which in English is unpronounced. This abstract 'invisible' complementizer can be called 'C'. The next step after we have built the IP similar to (3) is to merge in the complementizer:

(4)



Depending on whether this structure is a main or embedded clause, the inflectional element may or may not invert to the C-position in English. For the sake of simplicity suppose this is an embedded clause so that no auxiliary inversion takes place. However, the wh-object must still undergo movement. If this proceeds in the same way as subject movement, we will take this wh-element and merge it into the top of the structure, selecting the C to be the label and placing the wh-element in the specifier position:

(5)

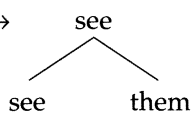


Again, movement is part of the structure-building process; every time a movement takes place the structure grows.

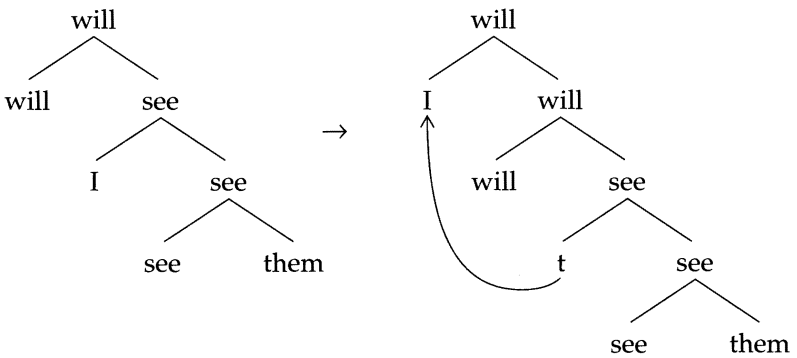
MOVEMENT AS STRUCTURE-BUILDING MERGER

Chomsky distinguishes two possible types of merger:
External Merge takes two independent elements, joins them in a single structure and selects one as the label:

see + them →



Internal Merge takes two elements, one included in the other, joins them in a single structure and selects one as the label:



EXERCISE 8.1

Given the Numeration (Q_1, may_1, $arrive_1$, who_1), where Q is an abstract interrogative complementizer, show the step-by-step convergent derivation using internal and external mergers for the sentence *who may arrive?*

8.2 The motivation for movement

Minimalism assumes no linguistic processes should exist beyond those required to satisfy the Bare Output Conditions imposed on the linguistic system by the interface systems at LF and Phonetic Form (PF). Clearly therefore some process which builds lexical elements into structures is necessary to produce expressions which are interpretable at both interfaces. You may be forgiven, however, for wondering why movement is necessary. Even if movements are seen as structure-building operations, structures can be built without them and so surely they are superfluous complications to the grammar.

Chomsky argues that this is not the way to view things at all (Chomsky, 2001b). Given a structure-building operation which takes two elements and combines them into a new object, there are two logically possible situations: the elements that are merged are independent of each other, in which case we are dealing with External Merge, or they are not, in which case we are dealing with Internal Merge (i.e. movement). A system which excludes one of these possibilities would be more complex precisely because it would exclude a logical possibility. Thus, Chomsky claims that the very existence of a Merge operation predicts that there will be movement. The exclusion of the possibility of Internal Merge would be an extra stipulation and hence non-minimal. 'Displacement is not an "imperfection" of language; its absence would be an imperfection' (Chomsky, 2001b, p. 8).

Of course, it is one thing to argue that movement is not an additional superfluity and quite another to argue that particular instances of movements in a given derivation are necessary, and therefore grammatical, steps in that derivation. For this, a different argument is needed to show that movement serves a purpose in a derivation and that a derivation would crash without it. To get some understanding of what is achieved by movement means reviewing some ideas that were first proposed at the end of the GB era, and were mentioned briefly in chapter 3. Recall that English main verbs do not appear to move out of the VP and that, unless there is an auxiliary verb that can move to the inflection position, the inflection moves to the verb. Such a downward movement, however, is problematic for two reasons. First, downward movements are rare: the vast majority of the movements considered so far are upwardly oriented. Thus inflection lowering is an exception in a system which prefers upward movement. Second, there is a reason why upward movements are preferred. It is only when an element moves upward that it can c-command and therefore govern its trace and hence conform to the Empty Category Principle (ECP). Lowering movements would leave behind an ungoverned trace, violating the ECP and ending up ungrammatical. In GB a number of possible solutions to this problem were suggested, such as Chomsky's own suggestion that there was a subsequent covert raising movement to set things right

before LF, at which the ECP applies (Chomsky, 1991a). However, given the possibility of covert movement, the simpler solution to the problem is that English main verbs do move to the inflection position, but covertly. Thus verb movement in English is like English quantifier raising and Chinese wh-movement, mentioned in chapter 4. French, on the other hand, has overt main verb movement.

If this is the case, the fact that English main verbs are inflected before they move leads to the conclusion that inflections are not inserted into the structure separate from the verbal element that bears them, but that verbs are inserted into a structure already in a fully inflected form. This raises two questions: what is the position that we have so far been referring to as I if it is not the position into which inflections are inserted? Why do finite verbal elements move there, overtly or covertly? To answer these points, Chomsky and Lasnik (1993) proposed that the I-position contains abstract inflectional features for tense and agreement which check that the inflections on the verb are the right ones to agree with the subject. So a third person singular subject will agree with the abstract inflectional features in I, of which it is the specifier. The verb, however, might have some other inflection as it does not enter into an agreement relationship with the subject when it is merged into the structure. The verb has to move to the I-position to check that it has the right agreement features. If it does, the sentence will be grammatical; if it does not, it will be ungrammatical.

The minimalist perspective recasts this idea in the following way. A verb is inserted into a structure from the Numeration along with its verbal features, which must be checked against the features of the abstract inflectional element by undergoing a movement. The movement places two elements in the relevant structural configuration for checking to take place. If the features check correctly, the derivation converges and a grammatical structure is derived; if not, the derivation crashes and the structure is deemed ungrammatical. Given that the cause of a derivation crashing is its illegibility at the interface levels, feature checking might be seen as a way of eliminating features that would otherwise be uninterpretable. Considering the inflectional features of the verb, it is intuitively obvious that the agreement features of person, number and gender appearing on the verb are not interpretable: the same features play a role in determining the meaning of a noun, but not of a verb. Hence checking the uninterpretable agreement features of the verb against the interpretable agreement features of the subject allows the verb's features to be *checked off*, i.e. deleted, and so prevented from causing a crash at LF.

A similar treatment might work for other movements, with the corresponding changes in assumptions. Nominal elements might, for example, be entered into a structure already fully inflected for Case, and hence they have Case features given to them in the Numeration rather than assigned them in certain structural positions in the syntax. From this perspective, Case-motivated movements, such as subject movement, raising and passivization (the A-movements), are movements required by the need for the Case features to be checked off. Case features are generally uninterpretable on the noun: they do not necessarily correspond to either thematic roles or grammatical functions – in a sentence like *we believe* ***him*** *to have hit Bill*, for instance, the accusative pronoun is an agentive subject. Thus a similar story to the checking off of the verb's agreement features might be told about Case-motivated movement. It seems reasonable that a nominative feature checks

against a feature possessed by the finite inflection and, given the connection between the presence of the light verb and the licensing of accusative Case, we might assume that accusative checks off against some feature of the light verb. Thus a basic account of Case phenomena from this perspective is that a nominative DP will move to the specifier of a finite IP to check its nominative feature, while an accusative DP may move (perhaps covertly) to the specifier of vP to check its accusative feature. Of course, the subject originates in the specifier of vP, but this is not a problem under a Bare Phrase Structure account, as anything merged after a head is first merged into a structure will be in specifier position and multiple specifiers are allowed.

Now consider wh-movement. The GB Theory assumed that the movement of a wh-element to the specifier of the complementizer entered both these elements into an 'agreement' relationship: a +wh-complementizer was assumed to agree obligatorily with a +wh-operator moved to its specifier position. Hence wh-movement is obligatory to the front of an interrogative clause. This 'agreement' relationship can be reinterpreted as a checking relationship: the moved wh-element checks its interrogative feature against an interrogative complementizer. There is a slight difference in this case, however, as both the features involved – the one on the wh-element and the one on the complementizer – appear to be interpretable. A wh-operator is presumably distinguished from a non-wh-operator precisely by its possession of an interrogative feature, while a clause is interpreted as interrogative if the head complementizer has an interrogative feature. In GB, both of these features were represented as [+wh], presumably signalling that they are the same thing. But this is not a necessary assumption: what makes a clause a question rather than a statement may not be the same thing as what makes an operator interrogative or non-interrogative. Let us suppose that different features are involved. Call the feature on the complementizer Q, and the one on the operator [+wh]. Now if a wh-element has both an interpretable [+wh] feature and a Q feature, which is not interpretable on the wh-element itself, then wh-movement is the same as the other movements that have been discussed: what motivates it is the need to check its uninterpretable Q feature.

From this perspective, all movements have a reason: things that move have an uninterpretable feature which must be checked off to prevent the derivation from crashing. 'Uninterpretable features are eliminated when they satisfy certain structural conditions: an uninterpretable feature of α must be in an appropriate relations to interpretable features of some β' (Chomsky, 2001b, p. 11). Movement appears to play a role in this checking procedure by placing the element with the uninterpretable feature in a position in which checking can take place. For phrases the relevant checking relationship appears to be the specifier of the head against which the features are checked, and for heads it appears to be a head adjunction position. These two positions can be referred to as the **checking domain** of a given head. The uninterpretable feature will check off only if it matches the features of the element that it is checked against, and hence a mismatch of features will cause the derivation to crash, as the uninterpretable feature will be unable to check off and will remain in the structure at LF, violating Full Interpretation.

This view of movement also offers an improvement over the GB version on grounds of simplicity. In GB, although no movement 'served a purpose' in the

way that movements do in the MP, different movements achieved grammatical structures for different reasons, as seen in chapter 3. Verb movement, for example, allowed a structure to be grammatical by getting the verb and the inflection together, hence satisfying the morphological requirements of the inflection. A-movements allowed DPs to achieve Case positions, hence satisfying the Case Filter, and Ā-movements seemed to satisfy certain semantic requirements placed on structures. In the MP, however, all movements have a single purpose: to check off uninterpretable features. Thus a more uniform theory of movement is achieved. Of course, one might wonder why natural languages contain uninterpretable features in the first place, which then necessitate movement, when a system which lacked uninterpretable features would not need to make use of movements at all and hence derivations could be made shorter and more minimal. This is a tricky issue which we would ideally like an answer to, but which remains a puzzle at the moment – the MP is after all always claimed to be a programme under development not a complete solution. However, the connection between the existence of uninterpretable features and the use of movements means that there is only one problem to solve: if we knew why natural languages contain uninterpretable features, we would have a ready explanation for why they utilize movements.

CHECKING THEORY

Uninterpretable features and checking: Elements may be inserted from the numeration bearing grammatical features which are uninterpretable at LF and which need to be eliminated for the derivation to converge. These features check against those of another element in the structure and, if the checked features match, the uninterpretable ones are **checked off**, allowing convergence.

Movement and checking: For checking to take place, the relevant structural relationship between the two elements must be established. Phrases check their features in the specifier of the relevant head while heads check their features by adjunction to the relevant head. It is movement that establishes these relationships. Movements are motivated, then, if, when they are carried out, an uninterpretable feature is checked off. In a minimalist system, this is the only motivation for movement and a movement that does not check off an uninterpretable feature will be superfluous and therefore ungrammatical.

EXERCISE 8.2 Considering the following movements, what uninterpretable features are checked? Are there any problematic cases for checking theory?

John$_1$ was examined t$_1$.
I wonder who$_1$ he is t$_1$.
They had$_1$ not t$_1$ read the contract.
Can$_1$ I t$_1$ help?

8.3 The nature of movement

So far movement has been characterized as a type of merger whose purpose is to check off uninterpretable features, but we have not gone very far into the details of movement itself. For example, in GB movements created traces in the positions vacated by the moved element. In the MP it is also assumed that movements leave traces behind, though this does not fall out from the assumption that movement is internal merger.

From the start of the MP, Chomsky has conceptualized movement in a different way from GB Theory. Of course, the term *movement* itself is a metaphor for what actually goes on in the biological process associated with linguistic knowledge and behaviour, about which we know next to nothing. Our views of linguistic movement are to help us understand how the linguistic system works at a certain level of abstraction. From this perspective one conceptualization may offer a better understanding than another. There are a number of ways to conceptualize syntactic movement. The most obvious, and the one mostly adopted in GB, was of a process similar to physical movement: an element occupies some location in a structure and the movement process 'picks it up' and deposits it in another location. However, empirical evidence for the existence of traces (as discussed in chapter 4, section 4.2.1) complicates this picture as physical movement does not involve anything being left behind in the extraction site. Thus the insertion of traces must be conceptualized as a separate, though perhaps simultaneous, aspect of movement. This view of the syntactic movement process can be called the **Move-Insert** version.

Another, quite different, conceptualization of movement is provided by the workings of a personal computer. Suppose we want to move a file stored on the hard disk to another location such as a CD or a pen drive. The process works like this. First the file is copied to the new location and then the original is deleted from its old location. The result is that the file has been moved from one place to another. Call this the **Copy-Delete** version of movement.

Can the Copy-Delete metaphor of movement provide us with more understanding of the syntactic process? One advantage of the Copy-Delete view is that some notion of a trace is built into it. Copying involves two elements: the source and the copy, both of which represent the same linguistic element in the same way that a moved element and its trace do. Furthermore, the source sits in exactly the position where the trace would be inserted. If the deletion of the source is partial, i.e. its phonological features alone, then the 'deleted' element has exactly the right characteristics to be viewed as a trace: it is a phonologically null version of the moved element. There may well be reasons why the phonological features of a copied element should undergo deletion, to do with the scattering of these features in an expression and the fact that phonology imposes a linear requirement on pronounced elements. Presumably elements should be pronounced in one place in the linear string. Semantics, however, does not require elements to be interpreted in one place: a wh-element, for example, may be interpreted as an operator in the position it is moved to, with its scope properties represented structurally from this position, and at the same time it may be interpreted as an argument of a

predicate in the position it moved *from*. Thus the deletion of the original element's semantic features would not only be unnecessary but would be harmful to the correct interpretation of the derived structure.

MOVEMENT AS COPY-DELETE

Movement involves the merging of a copy of some element contained in a structure with the structure itself. The original element then has its phonological features deleted and hence only the copy is pronounced. The deleted source element is what has been referred to previously as a trace and hence this notion comes 'built in' to this conception of movement.

One difference between the notion of an inserted trace of a moved element and that of a deleted source is that, while a trace is assumed to be a single empty element with no internal structure, the source has exactly the same internal structure as its copy. There is some evidence that traces should be viewed as structurally identical to the moved elements, however. Recall from chapter 4 the discussion of reconstruction which was used to account for examples such as the following:

(6) [Which picture of himself$_j$]$_i$ did John$_j$ display t$_i$?

The problem is that after movement the anaphoric reflexive is too high to be c-commanded by its antecedent and hence is predicted to be ungrammatical, contrary to fact. The proposed solution was to assume that most of the moved DP structure is carried along with the wh-determiner by a process termed pied-piping and that, given that only the determiner need be in its scope position for semantic purposes, the rest of the DP will be reconstructed back into its original position at LF:

(7) [Which t$_i$] did John$_j$ display [picture of himself$_j$]$_i$?

Ignoring the problem of the ungoverned trace within the wh-phrase, there is reason to believe that this is not a good solution. Consider a more complex example:

(8) [Which picture of himself$_j$]$_i$ did Bill$_j$ say [t$_i$ Sue liked t$_i$]?

The grammaticality of this clause cannot be accounted for by the assumption that the pied-piped material is reconstructed back into its original position, as in this position the reflexive is too far from its antecedent in the next clause up:

(9) [Which t$_k$]$_i$ did Bill$_j$ say [t$_i$ Sue liked [picture of himself$_j$]$_k$]?

Instead we would have to assume that the pied-piped material is reconstructed into the position of the intermediate trace:

(10) [Which t$_k$]$_i$ did Bill$_j$ say [[picture of himself$_j$]$_k$ Sue liked t$_i$]?

In this position it is perhaps possible for the reflexive to be properly bound. The problem is, however, there is no reason why the pied-piped material should be reconstructed into this position. It might make sense to assume that non-wh-material is put back into its original argument position where it is interpreted as the object of the predicate, but the position of the intermediate trace was a position that the moved wh-element occupied on its way to its scope position, forced by locality restrictions on movement. There is no interpretative reason why the wh-element occupied this position in the first place and hence no reason why the pied-piped material should be reconstructed back there.

The Copy-Delete view of movement provides us with a simple solution to these problems. As traces are identical to the moved element except that their phonological features are deleted, there is no need to resort to reconstruction. Consider what the structure of (6) would look like assuming a Copy-Delete version of movement:

(11) [Which picture of himself$_j$] did John$_j$ display [~~which picture of himself$_j$~~]?

Assuming that the semantic content of the source is not deleted, the contained reflexive is still present in the lower position from which it can refer to the subject. Moreover, intermediate traces will have exactly the same properties, as they are copies of the source which themselves have their phonological content deleted:

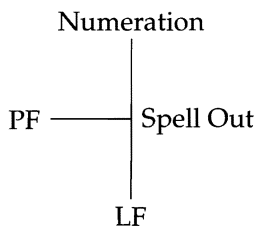
(12) [Which picture of himself$_j$] did Bill$_j$ say [[~~which picture of himself$_j$~~] Sue liked [~~which picture of himself$_j$~~]]?

Of course, other complications concerning the reconstruction analysis, such as the ungoverned trace left behind by the reconstructed element, are also addressed under these assumptions as no lowering movement takes place.

8.4 Overt and covert movement

The previous chapter characterized the computational process as a derivation from a Numeration to a fully built LF with a branching off to PF at a point called Spell Out, as in the following diagram:

(13)



All steps in the derivation, internal or external mergers, prior to Spell Out will have an effect on the possible shape of the PF, at least so far as hierarchical

structure imposes limits on linear order. In the post-Spell Out part of the derivation, however, any derivational step can have no effect on PF and hence will be covert. For obvious reasons, then, all overt elements in the Numeration must be merged into the structure before Spell Out, otherwise their phonological content will end up at LF and, not being semantically interpretable, cause a crash. Cases of internal merger, however, can take place before or after Spell Out as their function is to check off uninterpretable features and it doesn't matter when these are checked off, as long as it is before LF. We therefore expect to find instances of overt and covert movement.

The important question that remains to be answered is what makes the difference between overt and covert movement: why would an element move before or after Spell Out? Chomsky has experimented with a number of possible answers to this question, but one constant thread is that covert movement is preferred to overt movement and hence overt movement takes place only if forced. The first idea along these lines, dating back to the original work on the MP (Chomsky, 1993), was that the whole computational process obeyed a general principle called **Procrastinate**, which favoured the late application of any process. Scheduling a process late inevitably places it after Spell Out and hence covert operations conform to Procrastinate while overt ones do not. PF requirements might sanction the violation of Procrastinate. For example, the fact that a single structure must be assembled by PF, otherwise linearization could not take place, means that all relevant instances of external merger must take place before Spell Out, given that phonological processes cannot build structures. Overt movement, on the other hand, requires further motivation. Chomsky, following Pollock (1989), suggested that a distinction between **strong** and **weak** features might determine when movements are overt and covert. The suggestion is that some features are strong: these must be checked off before PF, where they are intolerable. Weak features, on the other hand, can survive at PF without causing a crash; hence these will only be checked off after Spell Out, in accordance with Procrastinate. Linguistic parameterization, according to this view, is a matter of whether a given feature is strong or weak in a language. If the language decides a feature is strong, this will be associated with overt movement of the element bearing the feature, whereas if the language opts for a weak feature, the language will display covert movement.

There are, however, a number of weaknesses in this view. The notion of strong and weak features is not particularly explanatory, as it simply boils down to a stipulation that some languages have overt movement while others have covert movement. There is a rather severe circularity here: we explain that an element moves overtly by assuming that it has a strong feature and we assume that it has a strong feature because it moves overtly. Certainly there seems to be no explanation why strong features should be problematic at PF specifically, as it simply is not the case that strong features are distinguishable from weak ones either phonetically or morphologically. Given that subject movement is overt in English, for example, subjects must bear strong features that objects do not. But subjects are generally not phonetically or morphologically distinct from objects in present-day English, having no Case-marking. Furthermore, even if the notion of Procrastinate does have an 'economy' flavour to it in claiming that some types of movements are preferable to others, this is not particularly a minimalist

assumption, as it imposes a condition on the computational procedure over and above those imposed by the Bare Output Conditions (i.e. Full Interpretation).

A second proposal that Chomsky put forward in 1996 ridded the system of the somewhat dubious role of PF in distinguishing strong and weak features and claimed that strong features were simply those that the system as a whole could not tolerate. Therefore, as soon as a strong feature is merged into a structure, the next step in the derivation must be to eliminate the feature if it is not to crash. Several positive consequences follow from this proposal. Consider the fact that the movement of verbs, subjects, wh-elements, etc. are typically related to functional elements: verbs move to I, subjects move to spec IP and wh-elements move to spec CP. As these functional elements are not merged into the structure at first, it follows that lexical elements cannot carry strong features, as they could not check these until after the relevant functional element is merged into the structure several steps later in the derivation. As a strong feature must be eliminated in the next step in the derivation, all derivations would crash if lexical elements had strong features. Therefore, Chomsky concluded, strong features must be associated with functional elements. Given that the strong/weak division is what characterizes linguistic parameterization under these assumptions, the conclusion that only functional elements can bear strong features jibes well with an assumption made in GB Theory, known as the Functional Parameterization Hypothesis. This has its roots in work by Borer (1984), who located all linguistic variation in the different behaviours of functional elements.

In addition to this view of the strong/weak distinction, Chomsky also attempted a more principled account of why covert movement should be preferred over overt movement, replacing Procrastinate. Overt and covert movement have always been assumed to be identical in that they both involve the movement of an element in its entirety. From an overt point of view, this makes sense. A structural element is typically viewed as a bundle of features of various types, including syntactic, semantic and phonological ones. Before Spell Out, this bundle is intact and clearly behaves as a unit in all pre-Spell Out processes. At Spell Out the bundle is divided and the phonologically relevant features are sent off towards PF. The question is, what happens to the remaining features after the phonological ones have been stripped away? Do they continue to behave as a unit or can specific processes manipulate subsets of these features? Given that post-Spell Out movement is invisible, it is difficult to say what actually goes on. However, the movement of the complete bundle of features that constitute a post-Spell Out element is non-minimal given that the motivation for the movement is to check off only a subset of them. Suppose, then, that minimal movements only move the subset of features actually involved in checking. This would be the norm in a minimal system as it involves less material undergoing movement. The question that now arises is why minimal movements do not take place before Spell Out: why does pre-Spell Out movement involve the entire feature bundle? It may be that this has a phonological motivation in that PF is a linearization of elements and therefore elements must be kept together so that the system knows where in the linear string they are to be realized, though obviously many additional assumptions are needed for this to add up to a complete story. Nonetheless, if this view can be sustained, we end up with an account for why post-Spell

Out movement is preferred to overt movement: post-Spell Out movement is movement of the checking features, while overt movement is movement of the entire element, and as such the former involves the movement of less material.

OVERT AND COVERT MOVEMENT

Overt movement in the MP is characterized as movement before Spell Out and **covert movement** as post-Spell Out movement.

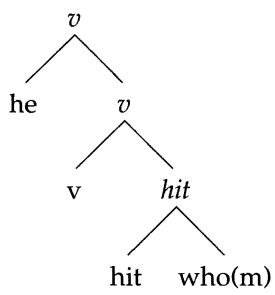
Procrastinate: A stipulation that all processes be delayed for as long as possible. Therefore post-Spell Out movement is preferred to overt movement. This principle was replaced in later versions of the minimalist grammar.

Strong and weak features: Strong features are associated with overt movement and weak features with covert movement. The original idea was that strong features must be checked off before Spell Out as they cannot be tolerated at PF. This idea was later refined under the assumption that strong features induce a movement as soon as they enter a structure, and as overt elements are merged into a structure before Spell Out then strong features will also be inserted into a structure before Spell Out. This will force movements to check strong features to be overt.

Feature movement: Covert movement, because it takes place after Spell Out and hence is not restricted by phonetic consequences, can be characterized as the minimal movement of the checking features themselves. Assuming this to be a more minimal movement than moving a whole category, this provides us with an explanation to why covert movement is preferred to overt movement.

At this point, it might be useful to look at a detailed example. Suppose that the derivation has reached the point at which all the arguments of a verb have been merged into the basic VP. In the Numeration there was an interrogative accusative element and a non-interrogative nominative element. The structure therefore stands as follows:

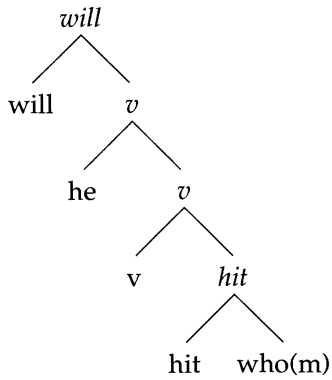
(14)



At this point the subject has an uninterpretable Case feature and the object has uninterpretable Case and Q features. None of these features is strong, however, as all elements that have been so far merged are thematic (with the possible

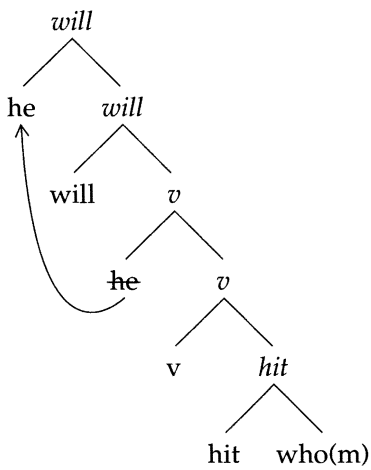
exception of the light verb). The next step in the derivation will be to merge in an inflectional element. Suppose the modal *will* is in the Numeration:

(15)



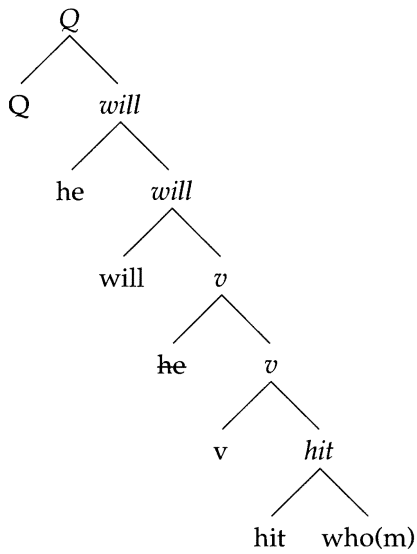
Apparently English inflections carry a strong feature which forces the subject to move to their specifiers. We will consider the nature of this feature in a little while, but for now let us assume that some feature of the inflection is strong and the next step in the derivation must therefore eliminate this feature. Thus the subject is copied to the spec IP position:

(16)



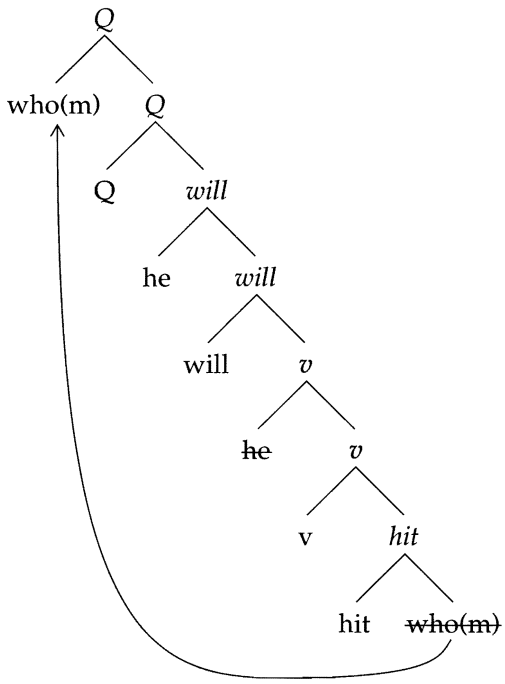
In this position not only can the strong feature of the inflection be checked off, but so too can the nominative feature of the subject. Note that this was not the reason why the movement took place at this point in the derivation, as Case features, being features on thematic elements, cannot be strong under the assumptions made so far. A feature which is checked under a movement motivated by the checking requirements of another feature is known as a **free rider**: they get a free ride on the moving element to their own checking positions. Finally an interrogative complementizer is merged into this structure. Let us represent this as Q:

(17)



English interrogative complementizers also have a strong feature, forcing a wh-element to move to its specifier. Again, we will consider the nature of this feature later. The next step must therefore be to eliminate this feature and to copy the wh-element to spec CP:

(18)



If this is an embedded clause, no other movement takes place within it and all uninterpretable features it contains are checked, though the issue of the accusative feature of the wh-element remains to be discussed in a subsequent section.

8.5 Properties of movement

So far we have demonstrated how Internal Merge works to achieve similar movements to those we introduced in the earlier chapters of this book. Till now there has been no evidence that the view of movement adopted in the MP has any advantages over the Move α of GB Theory.

8.5.1 The direction of movement

Recall that GB generally assumed that movement worked in an upward direction, although there may have been one or two exceptions. The explanation why movements are upward had to do with the ECP (see chapter 4, section 4.6.1): the trace of a lowered element would be higher than its antecedent and hence not properly governed. Of course, in the MP there is no ECP, as this would be an extra condition above those required by the Bare Output Conditions. Although the ECP was assumed to apply at LF, there is no semantically motivated reason why empty categories must be properly governed. One might hope, therefore, that the MP would have a better explanation for why movement is upward.

Movement is a kind of merger. So perhaps some property of Merge might explain why movements are always upwards. Consider how External Merge works. It is an operation which takes two independent elements, joins them together as a single element and selects one to be the label. The simplest way to conceptualize this process is to assume that the two merged elements are themselves unanalysed entities and as such the only way to merge them is to combine the two top nodes of each. This is not the only logical possibility, however, and we might conceive of mergers which merge one element into the middle of another. But, to go back to the metaphor of the model kit used in the previous chapter, this would be like taking an already constructed part of the model, breaking it in two at a certain place and gluing it back together again with a new piece inserted in the place where the break was made. This would not make much sense, as surely if we wanted to insert the new piece into this position, we would have done it at the relevant point, before the other part was complete. The same applies to External Merge. If we want to insert an element X into another element Y at some point P in the middle of Y, the easiest time to do it would be as we are building P. Therefore merging elements from their tops seems to be the minimal application of Merge. Chomsky has recently referred to this as the **No Tamper Condition** (Chomsky, 2004b): once a structure has been built, we do not tamper with the internal arrangement of that structure. Note that under present assumptions about movement, when an element is 'moved' out of a structure we do not tamper with the internal arrangement of that structure. All we do is make a copy of some element inside the structure and delete the phonological content of the original, leaving the structure untouched.

If merger always takes place from the top, it follows straightforwardly that internal merger can only merge the copied element with the top of the other structure and therefore that movement will always be upward. Consider what would

happen if an element were moved downward in a structure. We would make a copy of the element, deleting its phonological content, and merge the copy at some point lower than the deleted source. But any point lower than any other in a structure necessarily must be inside some structure and hence all lowering movements would violate the No Tamper Condition. It therefore follows on minimalist assumptions about the movement process that movements will always be to the top of a structure under construction, and as a result be in an upward direction.

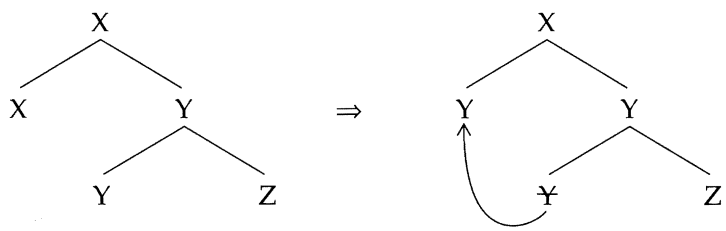
THE NO TAMPER CONDITION

The simplest and therefore minimal way to merge two elements is to merge them at their top nodes. The No Tamper Condition states that once a structure is built we cannot tamper with its internal arrangements and hence nothing can be merged internally to another element. This has the consequence that movements will be in an upward direction.

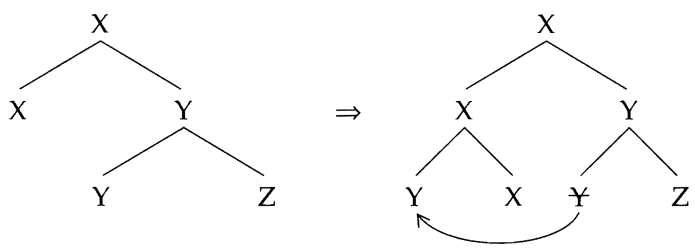
8.5.2 Head movement

The No Tamper Condition rules out not only lowering movements, but also upward movements not to the top of the structure under construction. This is problematic for some movements. For example, in the case of head movement, in GB it was assumed that when a head moves, it moves either into another head position or to adjoin to the other head. But neither of these options moves the head to the top of the structure, as the following diagrams demonstrate:

(19) a.



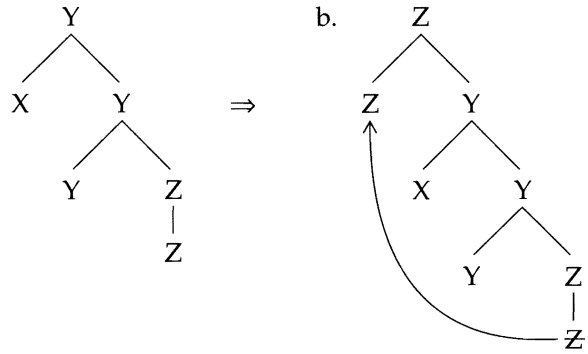
b.



While it is not entirely clear that the kind of substitution movement demonstrated in (19a) is valid under minimalist assumptions, what is clear is that neither (19a) or (19b) moves the head Y to the top of the structure. Therefore head movement is an exception to the No Tamper Condition: some part of the internal structure of the projection of X is changed when Y moves within this projection.

There are a number of possible ways out of this problem. One might be to deny that head movement works in the way assumed in (19). Following Surányi (2004) it might be proposed that a head does move to the top of a structure. The merger operation would have to select the moved head as the label of the structure to stop the moved head from being analysed as moving into specifier position, yielding the following analysis:

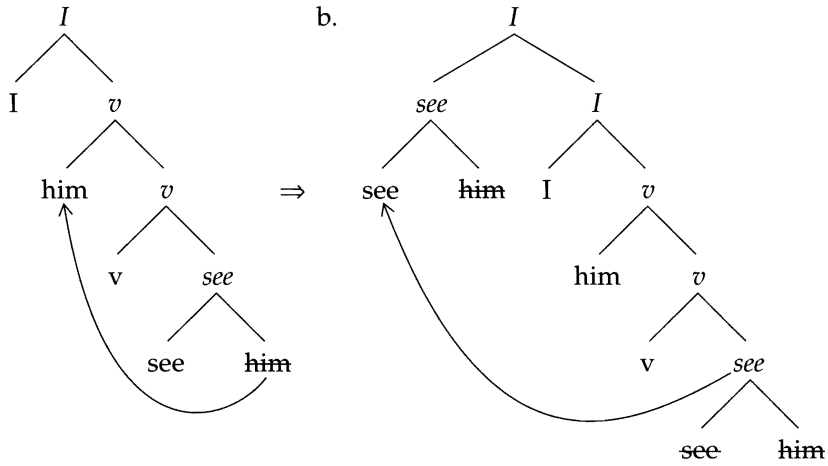
(20) a. b.



Chomsky, however, has generally denied that a moved element can be selected as the projecting head (Chomsky, 1995a), though his arguments for this assumption are based on phrasal movement and do not necessarily preclude Surányi's analysis.

Another possible solution would be to deny that there is any such thing as head movement at all. To account for phenomena that appear to involve a moving head, this would have to be reanalysed as phrasal movement. One way to achieve this is first to move all other elements out of a structure, leaving only the head, and then to move the phrase so that only the head appears to be moving. When a phrase that has had things moved out of it is itself moved, this is known as **remnant movement**. For example, suppose a language overtly moves its object out of the VP into the specifier of the light verb to check its accusative Case. This would leave the verb as the only element inside the VP. If we then move the VP, it would appear that only the verb was moving and specifically the object would appear to be left behind:

(21) a. b.



A third possibility is to assume that head movement is simply different from other movements and therefore does not obey the same conditions, such as the No Tamper Condition. As the No Tamper Condition, or indeed any condition on movement, falls out from minimalist assumptions concerning the movement operation, to claim that head movement is different from other movements is to claim that head movement is not movement per se. This is indeed what Chomsky claims, arguing that head movement may be more morphologically motivated and as such something which takes place on the path towards PF (Chomsky, 1995a). If this is so, then head movement is not a syntactic process at all and most of our previous assumptions about it will probably have to be reconsidered. Chomsky, however, has not gone into details concerning how a non-syntactic head movement would work and so at the moment it remains an unexplored possibility. The proposal is not without problems, however, not the least of which is that head movement shares many properties with other movement such as locality and upward direction. If head movement is not syntactic, then these similarities must be purely accidental.

HEAD MOVEMENT

Head movement appears to violate the No Tamper Condition. There are three approaches which aim to overcome this problem:

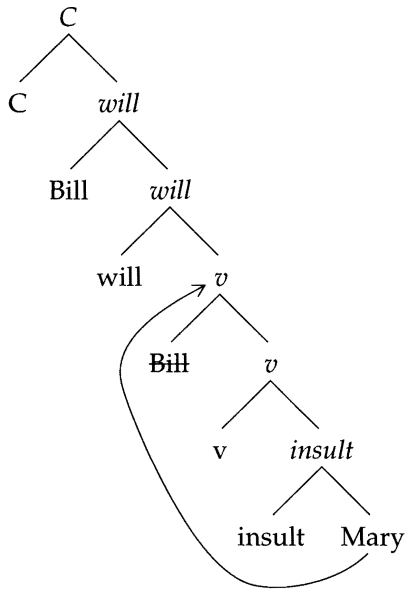
- Heads move to the top of the structure under construction and are selected as the label of the new construction.
- There is no head movement. What looks like head movement is really phrasal movement where all material apart from the head has already been moved out of it.
- Head movement is not syntactic movement but morphological/phonological in nature and as such happens on the line from Spell Out to PF.

8.5.3 Category and feature movement

A second problem arising from the No Tamper Condition concerns covert movement. Little has been said about this so far, precisely because the details are far from straightforward. Consider the simple case of the checking of the accusative feature of an object in English. As we have said, Case features are uninterpretable and therefore the object's accusative Case will have to be checked off if the derivation is not to crash. If the checking of the accusative is done in a similar way to the checking of the nominative, the object will undergo a movement to its checking position, indicated above to be the specifier of the light verb. However, in English this process is obviously covert as there is no reason to believe that the object overtly moves out of the VP. Covert movement, as we have said, takes place after Spell Out. But by the time the derivation has passed Spell Out the light verb is already included within a more fully built clause, with at least an IP and a CP

built on top of it. Covert movement to the specifier of vP would therefore seem to violate the No Tamper Condition:

(22)



Thus it seems that covert movement also does not conform to the No Tamper Condition, though the similarities between overt and covert movement, especially that the upward direction of covert movement would argue against this. If covert movement is characterized as feature movement, as discussed in the previous section, it may well not be identical to overt movement. For example, it may not be necessary to assume that features are moved to the specifier position for checking purposes. Perhaps feature movement is more like head movement, adjoining to the head that they check against. Given that we already know that head movement is not well behaved with respect to the No Tamper Condition, the association between head movement and (covert) feature movement may not be a bad move. Yet none of the suggested ways of treating head movement is suitable for treating covert movement: we cannot assume that features move after Spell Out to the top of the structure and are selected as the label for the new structure, as we don't want the features to move to the top of the structure but to somewhere in the middle of it; remnant movement is not an appropriate way to characterize feature movement, as it is unclear what the features would be a remnant of, and besides this also assumes that the remnant moves to the top of the structure. Finally Chomsky's assumption that head movement is not syntactic cannot be applied to feature movement as this is by definition a syntactic process.

One solution to these problems might be to abandon the idea that feature movement takes place after Spell Out. Given that feature movement does not pied-pipe all features, it would follow that, even if it took place before Spell Out, the phonological features would be left in their original position and hence the movement would remain invisible. This assumption would then allow covert movement to take place at an earlier step in the derivation and hence to the top of a structure

which has only reached the relevant stage. The object's accusative features, for example, can check off by merging with the structure headed by the light verb before the inflection is merged into the structure. Unfortunately these assumptions undermine the account of why covert movement is preferred and overt movement happens only when forced. Recall that overt movement was associated with a strong feature which, because it was inserted into a derivation before Spell Out, would force a pre-Spell Out movement, the assumption being that pre-Spell Out movements were overt. But if pre-Spell Out movements can be covert, there is no reason why the checking of strong features should be associated with overt movement, and indeed if feature movement is the minimal movement, it is difficult to know why overt movement should happen at all.

It is, in part, issues such as those discussed in this section which have led Chomsky to develop a new conception of how movement works. Chomsky's 'phases' approach to movement will be introduced after some more issues concerning movement phenomena.

For the time being let us put to one side the complexities of covert movement and continue with the assumption that overt movement is associated with strong features. We have yet to say exactly what these strong features are. Consider subject movement in English. As we know, the subject raises out of the VP into the specifier of IP where, if IP is finite, it checks its nominative Case. So subject movement has something to do with a strong nominative feature. But this feature cannot be the one on the subject itself as strong features can only appear on functional elements. Thus we suppose that a finite inflection carries a feature for checking the nominative Case and it is this feature that is strong. As soon as the inflection is inserted into a derivation, the subject movement will be triggered as the next step.

However, the following example presents a problem for this simple view:

(23) I believe [him to be worried].

In this example, the subject *him* has obviously undergone overt movement out of the VP into the specifier of the IP, as it is to the left of the infinitival inflectional element. Yet this subject bears accusative Case, which presumably is checked by covert movement to the specifier of the light verb above the Exceptional Case-Marking (ECM) verb *believe*. So why did the subject move to the specifier of the IP? It would appear that no feature was checked by this movement as the infinitival inflection carries no agreement features to check with the subject. But it is a central assumption of the MP that all operations happen for a reason and the sole reason for movement to happen is the checking off of uninterpretable features. We are therefore forced to assume that the inflection in (23) carries some uninterpretable feature which it checks off against the subject. What could this feature be? One possibility is that it is a newly discovered feature. But this would be highly circular, as the motivation for assuming the feature would be to account for the very same phenomena that the assumption attempts to explain. Thus we might hope to find independently motivated features which we might utilize in the account of this movement. One suggestion follows from the observation that IPs seem to require DPs in their specifier position. This goes back to the

start of GB Theory: clauses must have subjects – a principle known as the Extended Projection Principle (EPP: chapter 3, p. 87). As subjects tend to be DPs, it seems we can restate the EPP requirement that an I needs a DP specifier. This leads us to the conclusion that amongst other things, inflections check the categorial feature D. Suppose then that all inflectional elements have a feature which checks against a determiner (or a phrase headed by a determiner), which can be referred to as its checking D feature. Presumably this feature is uninterpretable and hence will need to be checked off for convergence. Moreover, in English the feature must be strong, triggering the overt movement of a DP to the specifier of the IP. Thus, in (23) the reason why the subject moves to the specifier of the IP is to check the inflection's strong D feature.

In recent times, Chomsky has generalized this idea and now uses it to account for why anything moves overtly (Chomsky, 2001b). Instead of assuming categorial checking features of one sort or another, Chomsky simply refers to this as the EPP feature, generalizing this notion far beyond the set of observations which motivated the principle in the first place. An EPP feature is simply a feature which requires something to move overtly to the specifier position of the element carrying it. So, in English, not only do inflections have an EPP feature but so do interrogative complementizers, which force a wh-element into the specifier of the CP. Of course there may be free riders which get checked along with the EPP feature and so overt movement does not only involve the checking of this feature. This goes some way towards characterizing the difference between overt and covert movement: overt movement is associated with the checking of an EPP feature and covert movement is not. This gets rid of the problematic association of overt movement with some PF requirement, though the assumption of an EPP feature is no less stipulatory than the assumption of strong features. It may be that the overt/covert distinction between movements and the linguistic variation associated with it has no independent explanation, and as such it stands as a small imperfection in the otherwise perfect system.

CATEGORIAL FEATURES AND EXTENDED PROJECTION PRINCIPLE FEATURES

Features which trigger overt movement do not seem to be related to Case or agreement features as some movements do not check these. The EPP of GB indicates that there is some requirement that a DP appear in the specifier of IP. This can be captured by assuming that inflections check the categorial feature D and that this is the strong feature that triggers overt subject movement.

This idea has been generalized to characterize all overt movement to check a strong feature of a functional head. This feature is known as the EPP feature.

8.5.4 Why do things fail to move?

If overt movement is triggered by the presence of an EPP feature on a functional element and features of Case and agreement, for example, are therefore always

free riders, one might wonder about what Case phenomena have to do with movement. Certainly, in GB Case Theory was central to the characterization of certain movements. Is this no longer the case? Consider the following observation:

(24) a. John seems [~~John~~ to be tall].
b. * John seems [~~John~~ is tall].

(24a) is the standard case of 'Case motivated' movement in GB: the subject of the non-finite clause must move to the subject of the raising verb in order to satisfy the Case Filter. The explanation for (24b) is different. In this example, the lower clause subject receives Case in the lower position and hence the Case Filter is already satisfied and has nothing more to contribute to the analysis. The cause of the ungrammaticality is that the trace is an anaphor and hence must be bound in its governing category, the lower finite clause. The antecedent lies outside of this clause and hence Principle A of the Binding Theory is violated. Let us now consider how the MP would approach these issues. In (24a), the subject *John* was originally merged into the VP and then was moved to the specifier of the non-finite clause to check the EPP feature of the inflection. At this point, no other feature was checked and the subject still had its nominative feature. At some later point the finite inflection is merged into the structure and this naturally has an EPP feature which needs to be checked by overt movement. Obviously the subject *John* is used to do this, demonstrating that one element can enter into a number of checking relationships involving the same feature: the subject checks the EPP feature of both inflections. As the upper inflection is finite, the nominative feature of the subject can check and hence the derivation converges.

What about sentence (24b)? From a minimalist point of view we would like the fact that here the nominative Case is already checked in the lower IP to account for why the subject cannot undergo a further movement. But this is not possible for two reasons. First, Case features are not responsible for overt movements in the first place, and hence whether or not they are checked should not affect whether the overt movement takes place. Second, we know that an element can enter into multiple checking relationships concerning the same feature, so even if the subject's features have all been checked, why would that matter?

To account for phenomena of this type, Chomsky therefore makes the following assumptions. An element can be in one of three states:

- none of its features has been checked
- some of its features have been checked, but not all
- all of its features have been checked.

For example, in sentence (24a), before the subject moves out of the VP, it is in the first state in which none of its features is checked. When it moves to the specifier of the non-finite inflection, it is in the second state, with whatever of its features that checks against the EPP feature of the inflection checked but the rest not. Finally when it moves to the specifier of the finite IP, it enters the last state, with all its features checked. The difference between (24a) and (24b) is simply that the subject enters into the last state, with all its features checked, when it

moves to the specifier of the embedded IP. Chomsky's assumption is therefore that when an element reaches this final state it becomes **inactive** and cannot enter into further mergers. Thus, as soon as all of the features of an element are checked it is as though that element becomes invisible to the system. Only elements with some unchecked features are **active** and are therefore visible to the system. The ungrammaticality of (24b) is thus due to the fact that the system would not be able to generate it, as *John* becomes inactive when it moves to the lower IP specifier position.

ACTIVE AND INACTIVE CATEGORIES

Active: An element is active if it still has unchecked features. This means that it can still enter into further syntactic processes and it can check features of other functional heads even if it has checked these features before.

Inactive: An element becomes inactive once all of its features have been checked and it cannot enter into further syntactic processes. This accounts for why a subject cannot be moved out of a finite clause, as its Case and agreement features will all be checked inside that clause and it will therefore become inactive.

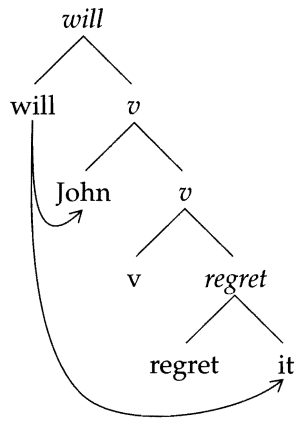
8.5.5 Locality

Finally in this section let us turn to a property of movement that has been the concern of syntax since the late 1960s: its locality. The notion of Minimal Link has already been mentioned as a condition placed on movements (chapter 7, p. 247). However, introduced into the system as a condition on movements, this is anything but minimalist, as it represents an extra restriction on the computation which is not understandable in terms of Bare Output Conditions. Perhaps the Minimal Link Condition might be somehow built into the movement process itself. One idea along these lines assumes that as movement is triggered by the insertion of an element which checks a given feature, it is this element that actually selects what will move to check the feature. Call an element which has a feature that needs checking a **probe**. Once a probe is merged into a structure it immediately looks for a relevant element to check this feature. The basic idea is that once a probe finds something that could check its feature it will look no further, and moreover it will find the nearest relevant element first. Hence only the nearest relevant element can possibly check the feature of a probe as further elements will never be considered. This will hold even if the nearest relevant element cannot check the feature, because it is inactive for example. In this case the feature will fail to be checked and the derivation will crash.

Let us take a basic example to show how this idea works. Consider the situation in which a transitive verb has had both of its arguments merged into the structure and a finite inflection has just been merged with the VP. The inflection

has an EPP feature to check and hence is a probe. There are two elements which could check this feature: the subject in the specifier of the light verb or the object in the complement of the verb. Which one of these will be copied and merged into the specifier of the IP to check the EPP feature?:

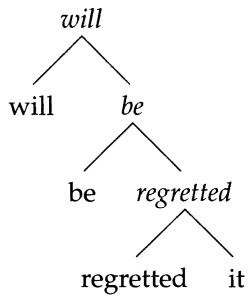
(25)



Clearly the subject is the nearer element and hence the first one the probe will select. The object will never come into consideration and hence only the subject will end up in the specifier of the IP.

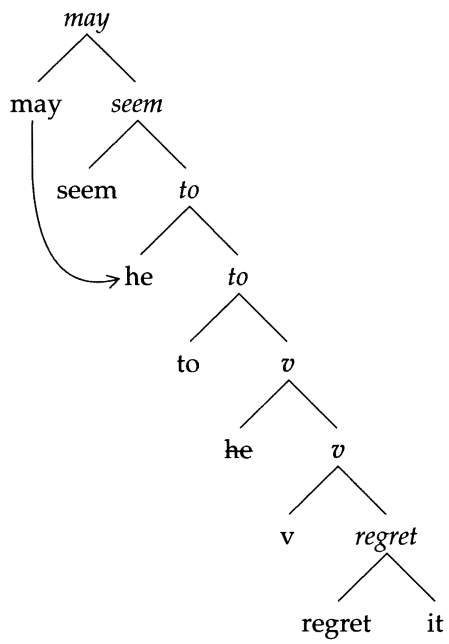
Now consider what happens in a passive, where there is no light verb and hence no agent or accusative Case checker. For the derivation to be able to converge, therefore, the object must have nominative features. After the finite inflection has been merged into the structure, the following situation arises:

(26)



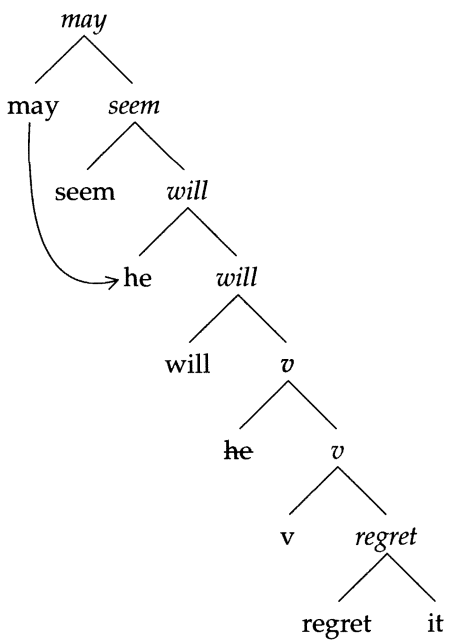
The inflection looks for the nearest element to check its EPP feature, and this time the object is the nearest as there is no subject. Thus the object will be copied and merged into the IP specifier. A similar situation arises with a raising verb. If the lower clause is non-finite, the subject will be merged into the specifier of IP to check the non-finite inflection's EPP feature. After this the raising verb will be merged into the structure and then its inflection. This also will be a probe and will look for the nearest element to check its EPP feature, namely the lower subject:

(27)



Finally, consider the situation when the lower clause is finite:

(28)

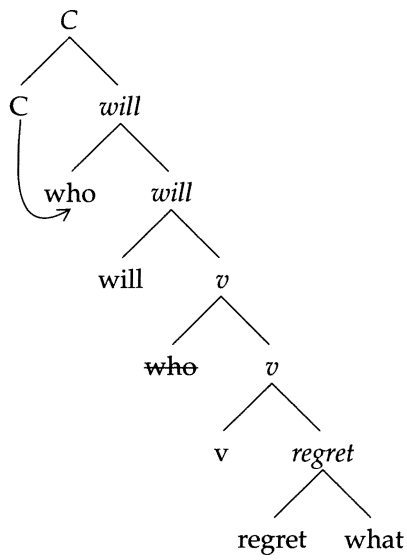


The nearest relevant element for checking the EPP feature of the upper finite inflection is the subject of the lower clause. However, as this has had all its

features checked in its present position, it will be inactive and hence will be unable to move further. The probe cannot look any further for something else to check its features and hence the derivation will crash, unless the Numeration contains an expletive *it* which can be merged into the IP specifier to check the EPP feature.

Essentially the same story can account for superiority and wh-island phenomena. With superiority the choice is between two wh-elements to check the EPP feature of an interrogative complementizer. After the complementizer has been merged into the structure, it looks for the nearest relevant element, as represented in (29):

(29)

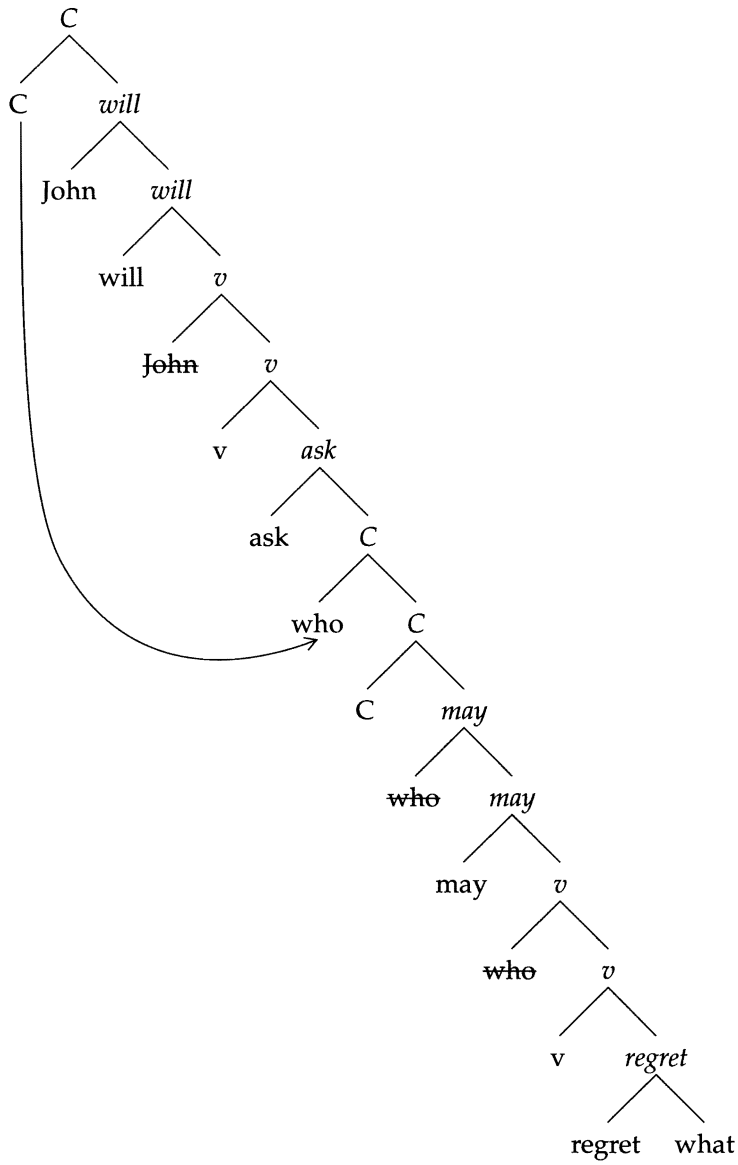


Clearly the superior (structurally higher) wh-element will be selected and hence this will be merged into the specifier of the CP. The object will never be selected, and so we account for the standard superiority observation that when there are two wh-elements in one clause only the superior one will move:

(30) a. Who will regret what?
b. * What will who regret?

While this situation is similar to the choice of the subject to check the inflection's EPP feature, wh-island phenomena are similar to A-movement out of a finite clause. With subject movement and superiority, there is a choice of two elements to move and the nearest one to the probe undergoes the movement. With A-movement out of a finite clause and wh-island phenomena, the nearest element to the probe cannot move as it is inactive, hence the derivation crashes. Consider the following structure:

(31)



Clearly the nearest wh-element to the interrogative C of the upper clause is the one in the specifier of the lower CP. However, this element has already moved to the specifier of the finite clause, where it had its Case and agreement features checked, and then it moved to the specifier of the lower interrogative complementizer, where it had its interrogative features checked. This element has had all its features checked and is therefore inactive and cannot move to check the EPP feature of the upper complementizer. Given that the complementizer cannot look further for another element to check its feature, the derivation will crash.

MINIMAL LINKS AND PROBES

The Minimal Link Condition: An original assumption of the MP to account for why movements are short is that short moves are preferred to long ones. Thus, given the choice of two possible elements to move, the nearest one to the landing site will be selected.

Probes: A probe is an element with an EPP feature to check. As soon as a probe is inserted into a structure, it looks for something to check this feature. The search will find the nearest element first and, after it has been found, the search is called off and no other possibility is considered. If the discovered element can move, it will, and the derivation moves one step closer to convergence; if not, the derivation crashes.

Although this takes us some way towards accounting for these locality phenomena, certain conceptual and empirical problems remain unaddressed. The conceptual problem is that the assumption that a probe looks for the nearest element to check its feature simply restates the Minimal Link Condition in terms of the structure-building process itself. There is, however, no good reason why the structure-building process should have the property in terms of Bare Output Conditions: what bearing on the interpretability of a structure could there possibly be if a probe were allowed to look further than the nearest relevant element to check its features? The empirical problem concerns long-distance wh-movement. It has been a standard assumption since the early 1970s that when a wh-element undergoes a long-distance movement, it does so in a series of short hops from one CP specifier to another. This assumption is based on empirical evidence from languages such as Irish in which complementizers 'agree' with a wh-element that has passed through their specifier position. In cases of long-distance wh-movement, all the complementizers below the final landing site of the wh-element take on a special form, indicating that it must have passed through all of their specifiers (McCloskey, 2000):

(32) an t-ainm a hinnseadh dúinn a bhí ar an áit
(the name C_{+wh} was-told to-us C_{+wh} was on the place)
the name that we were told was on the place

How can the cyclicity of long-distance wh-movement be achieved under the assumptions made above? First recall that what motivates an overt movement is the presence of an EPP feature on the probe, in this case a complementizer. So, for a wh-element to pass through a series of CP specifiers, we must assume that each complementizer has an EPP feature. This will work, as the lower complementizers will all be declarative and so will not check the interrogative feature of the wh-element. Hence this will have at least one feature unchecked and will remain active until it reaches its final destination:

(33) Who $C_{[+wh]}$ (did) John think [~~who~~ $C_{[-wh]}$ Mary said [~~who~~ $C_{[-wh]}$ Bill knows ~~who~~]]?

The problem, however, is explaining why each of the lower complementizers has an EPP feature. Declarative complementizers in English do not normally have an EPP feature, though they do perhaps in V2 languages (see chapter 4, section 4.1.3); hence there is no requirement that any element sits in the specifier of a declarative CP. However, in order to account for the cyclical movement in (33) we must assume that declarative complementizers *can* have EPP features. Thus, complementizers may optionally have an EPP feature. This will not affect the basic assumptions. An interrogative complementizer without an EPP feature will not attract a wh-element to its specifier, and hence its [+wh] feature will be unchecked, causing a crash; a declarative complementizer with an EPP feature will not be able to check a wh-element's features, if one is present in the structure, or, as seems more likely, there will be no wh-element at all and hence the EPP feature will not be checked and again the derivation will crash. But, if complementizers have optional EPP features, we cannot explain why each of the complementizers in (33) *must* have an EPP feature. Note that the derivation will still converge if only the top complementizer has an EPP feature:

(34) Who $C_{[+wh]}$ (did) John think [$C_{[-wh]}$ Mary said [$C_{[-wh]}$ Bill knows ~~who~~]]?

At the point where the interrogative complementizer is merged into the structure, the nearest wh-element is the one in the object position of the lowest clause and, as this has had none of its features checked, it is still active and hence able to merge at the top of the structure. Essentially, then, we are unable to account for the basic Subjacency conditions under these assumptions. The next section introduces Chomsky's latest proposals concerning the structure of the grammar, in part motivated by these problems.

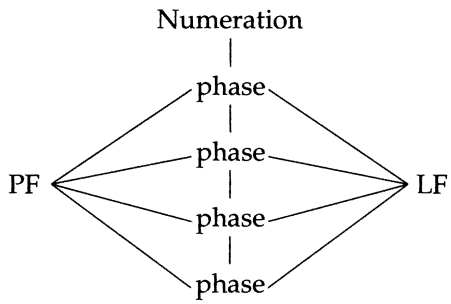
8.6 Phases

A recurrent notion concerning locality in movement phenomena since Ross's original work on islands (1967) is that some domains are opaque and do not let elements move out of them. Typical of this is the fact that subjects can move out of a non-finite clause, but not out of a finite one. On standard assumptions that the non-finite clauses that allow extraction from them are IPs whereas finite clauses are always CPs, this reduces to the statement that elements can move out of IPs, but not out of CPs. In recent work, Chomsky (2000c, 2001a) has captured this observation through the assumption that, at certain points in the derivation, part of the structure under construction becomes fixed and is unable to be manipulated further. Thus, if an element has not been moved by that point in the derivation, it cannot be moved after it. These points Chomsky refers to as **phases**. 'The computation maps LA [a lexical array = Numeration] to <PHON, SEM> [the pair of representations interpreted phonetically and semantically] piece-by-piece cyclically. . . . Call the relevant units "phases"' (Chomsky, 2001b, p. 4).

To accommodate the notion of a phase, Chomsky suggests a change to the basic working of the computation. Instead of having a single point at which a

derivation is 'spelled out', i.e. the derivation splits into PF- and LF-relevant operations, as soon as certain parts of a structure are complete, they are sent off to the interface components for interpretation, effectively fixing them at that point. The derivation may continue to build structures on top of the fixed part of the structure until another phase is complete and this will then be sent off to the interface components, continuing until the derivation is complete. This new conception of the computational procedure can be represented in the following way:

(35)



From this perspective, there is no single point for Spell Out, but multiple points at which a structure is 'spelled out' in terms of both phonetic and semantic interpretation. As always, this needs to be seen as a model of competence, not performance, and as such this does not model how people process speech. The point is not that people 'produce' language in small chunks, but that the grammar processes structures in chunks, thereby preventing certain grammatical operations from applying between certain parts of a structure.

A number of questions arise at this point about the details of this proposal. What, for example, counts as a phase and what does not? Moreover, how is long-distance movement possible if a structure becomes fixed at each phase? To answer the first question, Chomsky claims two particular phrases constitute complete phases in the derivation: the vP, headed by the light verb, and the CP. Empirically, it is obvious why CP is chosen as a phase, as this corresponds to an often opaque node for movement in many of the past approaches to locality, such as Islands, Subjacency and Barriers. In Chomsky's *Barriers* framework (Chomsky, 1986b), the VP was also identified as a potential barrier, carried forward to the assumption that vP constitutes a phase. Conceptually Chomsky claims that vP and CP have in common the fact that they represent certain semantically well-defined parts of a structure. The vP includes a predicate and all its arguments, including the external argument, and hence represents the basic proposition expressed by a clause. The CP is the final node of the clause including all other aspects of meaning added to the basic proposition, such as tense, mood and force. 'Ideally, phases should have a natural characterization . . . : they should be semantically and phonologically coherent and independent. At SEM [the semantically relevant representation], vP and CP (but not TP) are propositional constructions. . . . At PHON [the phonologically relevant representation] these categories are relatively isolable' (Chomsky, 2001b, p. 22).

As to the second question, Chomsky reinstates an old idea concerning how elements can escape from opaque domains. A relevant observation is that while

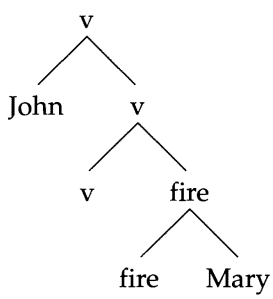
A-movements are restricted to tensed clauses, Ā-movements, such as wh-movement, can escape a tensed clause. The difference is that only the latter makes use of the specifier of CP and this seems to be an 'escape hatch' for movement. Recasting this idea in terms of phases, Chomsky claims that when a phase is complete, it is not the whole phase that is sent off for interpretation: the left edge, including the specifiers and the head, is not spelled out until the next phase is complete:

> Consider a typical phase (5), with H its head:
> (5) PH = [α [H β]]
> Call α-H the *edge* of PH. . . . A natural condition . . . is that β must be spelled out at PH, but not the edge: that allows for head-raising, raising of Predicate-internal subject to SPEC-T, and an 'escape hatch' for successive cyclic movement through the edge. (Chomsky, 2001b, p. 5)

Thus, at each phase only the complement of the head of the phase is spelled out, except for the last phase when everything is finally sent off for interpretation.

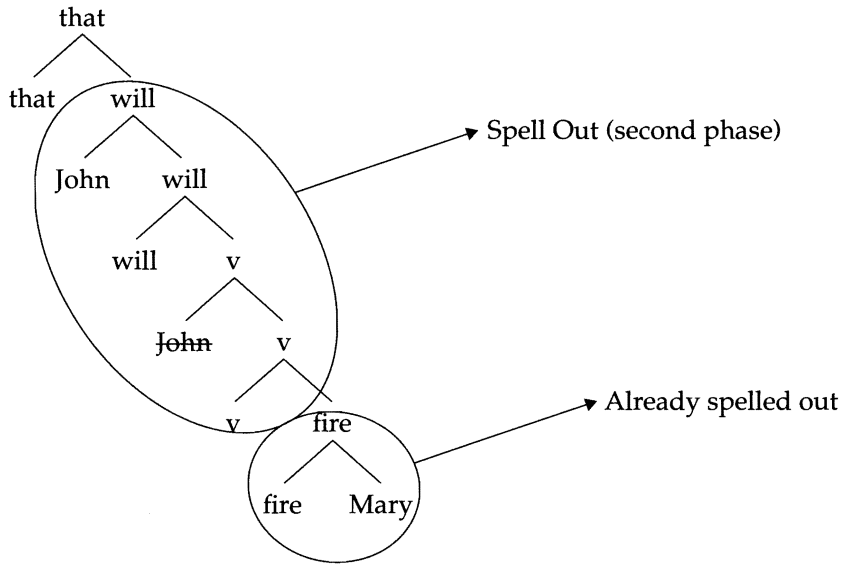
A simple example will serve to demonstrate these ideas. Let us consider a step-by-step derivation of a sentence such as *Bill might think that John will fire Mary*. The first step is to merge the lower verb and the internal argument to form the VP, and then in turn the light verb and the external argument to form the vP:

(36)



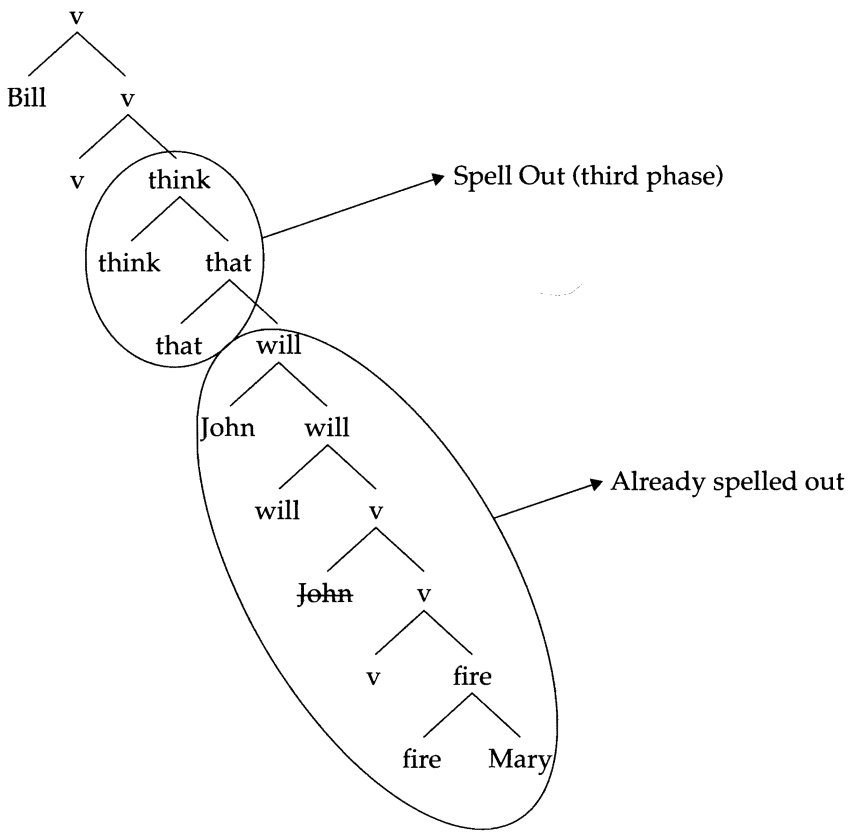
The light verb is assumed to have no EPP feature to attract the object to its specifier position (we will discuss 'covert movement' a little later) and hence at this point the vP is complete and the phase can be spelled out. This will mean sending the complement of the head of the phase to LF and PF for interpretation. The main verb and its complement are therefore fixed in their current positions and cannot undergo any further processing. The subject and the light verb, being at the left edge of the phase, are still syntactically active. The derivation continues by the addition of the inflection and the movement of the subject to check its EPP feature, the features of the subject also being checked off at this point. Note that only the subject is available for this movement as the object has already been spelled out in object position. So for this phenomenon at least, there is no need to resort to extra stipulations concerning a probe looking to the nearest possible element to check features off against. The complementizer is then merged and, as it has no EPP feature, it does not trigger any movement to its specifier. This ends the second phase and the complement of the complementizer, the IP, is then spelled out, fixing the subject and inflection in place:

(37)



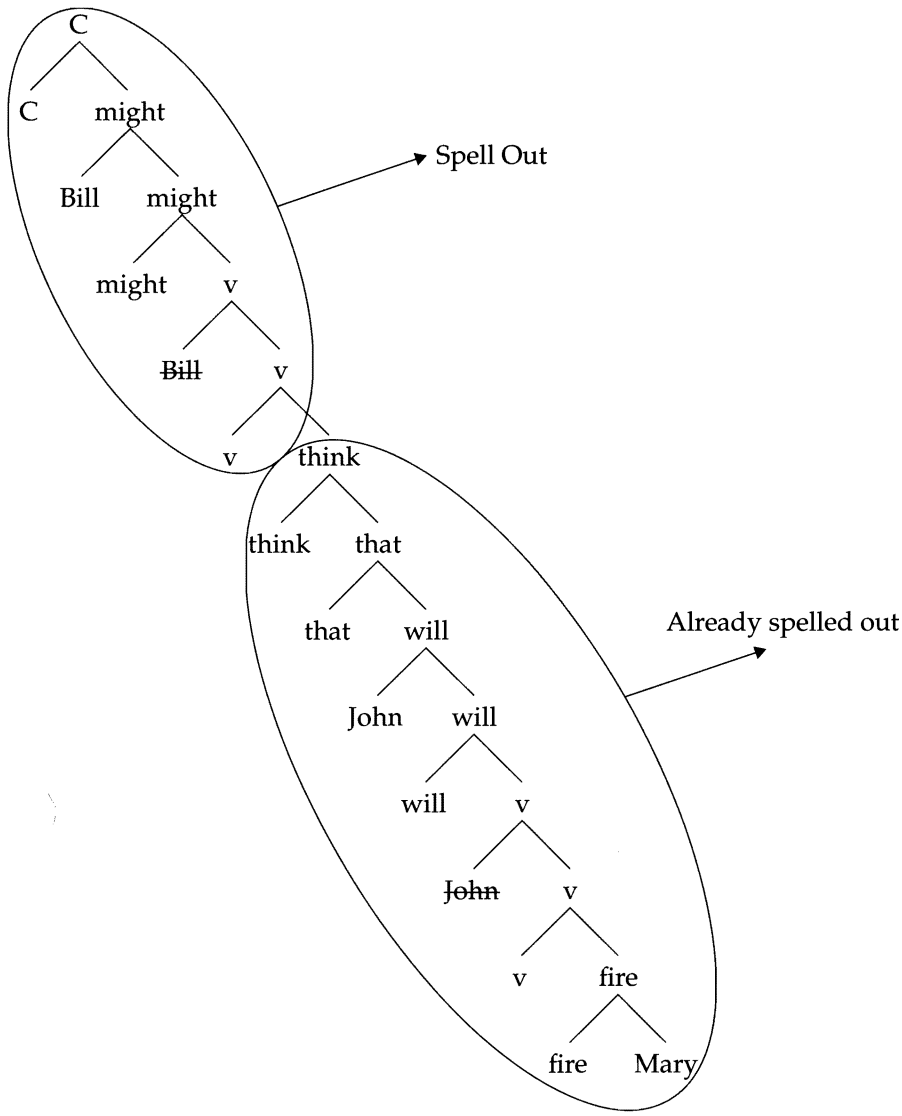
The derivation now continues to the next clause with the verb *think*, its light verb and subject being merged, in exactly the same way as the first phase. Once this vP is complete, the upper VP will be spelled out; the light verb and the subject, being at the edge of the vP, are unfixed at this point:

(38)



In the final phase the inflection of the matrix (main) clause is merged, the subject of *think* is moved to the specifier of IP and the abstract matrix complementizer is merged in. As this is the final phase, all that has not yet been spelled out will be and the derivation is complete:

(39)



Now let us consider the basic points of passivization, raising and wh-movement within this framework. In passivization, there is no light verb above the VP and hence the first phase will be the CP above it. For this reason, the object will not be fixed in place before the inflection is merged and can therefore move to check the inflection's EPP feature:

(40) $[_{CP}\ \underbrace{C\ [_{IP}\ \text{Bill was}\ [_{VP}\ \text{forgiven}\ \cancel{\text{Bill}}]]]}_{\text{First phase}}$.

Raising involves a subject of a non-finite IP that does not extend to a CP. For this reason the subject can raise out of this clause. Moreover, as the raising verb has no external argument, it has no light verb and therefore the next phase boundary will be the entire CP. This means that the lower subject will still be unfixed when the inflection is merged with the VP headed by the raising verb, and hence it can raise to check the EPP feature of the inflection:

(41) $[_{CP}\ \underbrace{C\ [_{IP}\ \text{Bill might}\ [_{VP}\ \text{seem}\ [_{IP}\ \cancel{\text{Bill}}\ \text{to}}_{\text{Second phase}}\ [_{VP}\ \underbrace{\cancel{\text{Bill}}\ v\ [_{VP}\ \text{forgive John}]]]]]]}_{\text{First phase}}.$

Finally, with wh-movement out of a finite clause, as the wh-element will move to the specifier of the lower CP first, it will not be fixed in position when this phase is spelled out. This allows it to move to the next clause, first to the specifier of the vP and finally to the specifier of the upper CP, where it will eventually be spelled out:

(42) $[_{CP}\underbrace{\text{Who C (did)}\ [_{IP}\text{John I}\ [_{VP}\cancel{\text{who John}}}_{\text{Fourth phase}}\ \underbrace{v\ [_{VP}\text{think}\ [_{CP}\cancel{\text{who}}}_{\text{Third phase}}\ \underbrace{C\ [_{IP}\cancel{\text{who}}\ I\ [_{VP}\cancel{\text{who}}}_{\text{Second phase}}$

$\underbrace{v\ \text{left}}_{\text{First phase}}]]]]]]]?$

Within a system where there are multiple points of Spell Out, it doesn't make much sense to talk about pre- and post-Spell Out movement; for this reason the original distinction between overt and covert movement must be dropped once phases are adopted. The conception of covert movement as feature movement could be adopted, leaving the problem of explaining why in some cases whole categories move and in others only the checking features do. Not only is the assumption that the EPP feature triggers category movement while its absence licenses feature movement ad hoc, but there isn't even any intuitive connection between an EPP feature and overt movement that could motivate the distinction.

Within the phases model Chomsky offers another view. Reverting to the idea that movement is category movement, he claims that covert 'movement' is not movement at all but long-distance checking. In other words, features may be checked between two elements in a structure without movement of any kind taking place. This has two positive consequences in a phase-based system. The first is that it immediately captures the notion of why covert movement is preferred to overt movement on straightforward minimalist grounds: overt movement always involves an extra step in the derivation, movement + checking, whereas covert movement involves just checking. Secondly, it allows the checking of features of a currently merged element against those of some element that has already been spelled out. This would not be possible if it were assumed that some features of an already spelled-out element could undergo actual movement, as any process that affects something that has already been spelled out is not possible: once spelled out, all parts of an element are fixed and cannot be altered. However, this does not prevent the features of a spelled-out element being used to check off uninterpretable features of an element in a current phase. An example

of where such a thing would be useful is in cases of 'long-distance wh-movement' in a language like Chinese where wh-elements do not move overtly. Obviously in Chinese a wh-element is spelled out in its original merger site, which may be any number of clauses away from the CP from which its scope is interpreted. Thus, the wh-element is spelled out long before the interrogative complementizer which it is associated with is merged into the structure. A connection between the two can be established, however, if we assume that the complementizer has a [+wh] feature which it checks off over a long distance against the already spelled-out wh-element. Obviously this would not involve the checking off of any feature of the wh-element itself, as if this had any unchecked uninterpretable features the derivation would have crashed in the phase that introduced the wh-element.

PHASES

Instead of the assumption that there is one point of Spell Out, it can be assumed that the computation proceeds in chunks, at the end of which the completed portion of the structure is sent off for interpretation. These chunks are called **phases**.

Phases: Relevant phases are assumed to be vP and CP. These correspond to parts of the structure which pertain to salient semantic aspects of the clause. The vP expresses the basic proposition and the CP adds all other aspects of meaning. The vP and CP also correspond to well-noted opaque domains for movement.

Escape hatch: To allow long-distance movement, it is assumed that not everything in a phase is spelled out. Specifically, the left edge of the phase remains uninterpreted until the next phase is complete. Thus an element moved to the specifier of the head of a phase will be able to continue to move.

Covert movement: As the distinction between pre- and post-Spell Out movement is confused in a model where there are multiple spell-out points, a new conception of the distinction between overt and covert movement is needed. It is assumed that only overt movement is actual movement (triggered by an EPP feature) and covert movement does not actually involve movement at all. Instead, features are allowed to be checked over long distances, providing that no unchecked uninterpretable feature is sent off for interpretation at the end of a phase.

The system is still not without problems, however, and there are many issues to be resolved. To discuss just one of these, consider a simple case of wh-movement from object position. In the first phase the main verb and the wh-object are merged to form the VP, and then the light verb and the external argument are merged into the structure. Before the phase ends and the VP is spelled out the wh-element must be moved to the specifier of the vP so that it remains available for movement in the next phase. But why would the wh-element move to the specifier of vP? If the language has overt object movement, there is no problem,

as we can assume that the light verb has an obligatory EPP feature and therefore the wh-element will move to check this. But, in a language like English, objects do not normally undergo overt movement and hence we cannot assume an obligatory EPP feature on the light verb. Chomsky assumes in this case that there is an optional EPP feature (2001a). If the light verb has this feature the wh-element will move to check it and the derivation can continue. If the light verb fails to have the feature the derivation will crash, as the wh-element will be unable to check its uninterpretable Q feature.

While this gives us an account of wh-movement, it raises a problem for non-wh-objects. If light verbs can have an EPP feature, why do we not get optional object movement? If the light verb is assigned an EPP feature with a non-wh-object, this object should move to the specifier of vP to check it and, if there is no EPP feature, the object will check its accusative features against the light verb without movement. Both derivations will converge and hence object movement should be optional. While there appears to be a 'least effort' effect here – for a wh-object only the derivation that involves movement can converge, whereas for a non-wh-object either moving or not moving is possible and hence not moving is preferable – Chomsky has been moving away from such explanations of ungrammaticality due to the complexities they cause for the computational system. Moreover, in this particular case it is not clear how to make the two derivations compete against each other, as they are related to different Numerations: one in which the light verb has an EPP feature and one in which it does not. This problem reoccurs for all cases of wh-movement from spec CP to spec CP, as all of these will have to go through the intervening spec vP to escape that phase.

8.7 Conclusion

If one compares the progress of the MP to that of GB Theory, one of the most striking observations is how much one version of the minimalist grammar differs from another. In GB, certainly new ideas were constantly being introduced and old ones modified or replaced, as is inevitable in scientific investigations. But in the MP it is not just the details of well-established principles that are adjusted; large assumptions about how the system actually operates are completely altered. It isn't even true to say that notions about what counts as minimal have remained stable over the period that the MP has been developing: the notion of competition, or economy, has been given a very much lesser role in later developments of the Program. It is understandable that non-linguist members of the scientific community with an interest in language, for whom this book is mainly intended, might not like this situation: how can linguistic theories be applied in other areas if they change so radically every two or three years?

The job of the linguist is to come to some understanding of how the linguistic system works. This can only be done by formulating questions about the system based on what we already understand and investigating these questions through the development of specific hypotheses. The MP has set itself a tough job; one in which it is not at all clear what the questions are, let alone the answers. It may

well be, therefore, that the MP has got ahead of itself and its questions are premature. However, to the extent that the questions can be posed and that possible answers to them are conceivable, progress is being made towards the goal of greater understanding of the system. The conception of movement as an instance of the application of the basic structure-building process, for example, has advanced our understanding of why the linguistic system has such a bizarre operation.

Moreover, even if the MP can fulfil its aims only in part, it will be capable of providing a depth of understanding that goes well beyond anything achieved so far, maybe even deeper than our current understanding of other natural systems in the universe. Although this note is rather promissory, the potential gains it offers make the attempt worthwhile.

Discussion topics

1 A quick comparison of chapters 3 and 4 with chapters 7 and 8 will reveal topics covered in the GB framework that have not been mentioned from the minimalist perspective. Which of these would you consider to be easily incorporated within a minimalist approach and which would be problematic?
2 Is locality an entirely expected property of a minimalist grammar?
3 The notions of economy and competition were central to the MP when first conceived. How far have they been removed from current versions?
4 In GB Theory, Case Theory played a central role in the analysis of certain movements. Do Case phenomena play any essential role in the minimalist analysis of movement?
5 Is the notion of Full Interpretation within GB Theory the same as that used in the MP?
6 How surprising would it be for Chomsky's intuition that the human language faculty is perfect to be found to be true?

References

Abney, S. P. (1987), The English Noun Phrase in its sentential aspect. PhD, MIT.

Adger, D. (2003), *Core Syntax: A Minimalist Approach*. Oxford: Oxford University Press.

Atkinson, M. (1992), *Children's Syntax*. Oxford: Blackwell.

Baker, C. L. (1979), Syntactic theory and the projection problem. *Linguistic Inquiry*, 10/4, 533–81.

Baker, M. (1988), *Incorporation: A Theory of Grammatical Function Changing*. Chicago: University of Chicago Press.

Baker, P. and Eversley, J. (eds.) (2000), *Multilingual Capital: The Languages of London's Schoolchildren and their Relevance to Economic, Social and Educational Policies*. London: Battlebridge.

Balcom, P. (2003), Cross-linguistic influence of L2 English on middle constructions in L1 French. In V. J. Cook (ed.), *L2 Effects on the L1*. Clevedon: Multilingual Matters, 168–92.

Bald, W-D., Cobb, D. and Schwarz, A. (1986), *Active Grammar*. Harlow: Longman.

Bellugi, U. and Brown, R. (eds.) (1964), *The Acquisition of Language*. Monographs of the Society for Research in Child Development, 29, 92.

Berwick, R. C. (1985), *The Acquisition of Syntactic Knowledge*. Cambridge, Mass.: MIT Press.

Berwick, R. C. and Weinberg, A. S. (1984), *The Grammatical Basis of Linguistic Performance*. Cambridge, Mass.: MIT Press.

Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999), *Longman Grammar of Spoken and Written English*. London: Longman.

Bickerton, D. (1984), The language bioprogram hypothesis. *Behavioural and Brain Sciences*, 7, 173–88.

Birdsong, D. (1992), Ultimate attainment in second language acquisition. *Language*, 68, 706–55.

Bley-Vroman, R. (1989), The logical problem of second language learning. In S. Gass and J. Schachter (eds.), *Linguistic Perspectives on Second Language Acquisition*. Cambridge: Cambridge University Press, 41–68.

Bloomfield, L. (1933), *Language*. New York: Holt.

Borer, H. (1984), *Parametric Syntax*. Dordrecht, Foris.

Borer, H. and Rohrbacher, B. (2002), Minding the absent: arguments for the full competence hypothesis. *Language Acquisition*, 10, 2, 123–75.

Bosewitz, R. (1987), *Penguin Students' Grammar of English*. Harmondsworth: Penguin.

Botha, R. P. (1989), *Challenging Chomsky: The Generative Garden Game*. Oxford: Blackwell.

Braine, M. (1963), The ontogeny of English phrase structure: the first phase. *Language*, 39, 1–13.

Braine, M. D. S. (1971), On two types of models of the internalisation of grammars. In D. I. Slobin (ed.), *The Ontogenesis of Grammar*. New York: Academic Press, 153–86.

Brown, R. (1973), *A First Language: The Early Stages*. London: Allen and Unwin.

Brown, R. and Hanlon, C. (1970), Derivational complexity and the order of acquisition in child speech. In J. R. Hayes (ed.), *Cognition and the Development of Language*. New York: Wiley, 155–207.

Bruner, J. (1983), *Child's Talk*. Oxford: Oxford University Press.

Burzio, L. (1986), *Italian Syntax*. Dordrecht: Reidel.

Carroll, S. and Swain, M. (1993), Explicit and implicit negative feedback: an empirical study of the learning of linguistic generalisations. *Studies in Second Language Acquisition*, 15, 3, 357–86.

Cazden, C. R. (1972), *Child Language and Education*. New York: Holt, Rinehart and Winston.

Chomsky, N. (1957), *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. (1959), Review of B. F. Skinner *Verbal Behavior*. *Language*, 35, 26–58.

Chomsky, N. (1964), *Current Issues in Linguistic Theory*. The Hague: Mouton.

Chomsky, N. (1965), *Aspects of the Theory of Syntax*. Cambridge, Mass.: MIT Press.

Chomsky, N. (1966a), *Topics in the Theory of Generative Grammar*. The Hague: Mouton.

Chomsky, N. (1966b), Linguistic theory. In R. Mead (ed.), *North-East Conference on the Teaching of Foreign Languages*. Reprinted in J. P. B. Allen and P. van Buren (eds.) (1971), *Chomsky: Selected Readings*. Oxford: Oxford University Press.

Chomsky, N. (1969), Linguistics and philosophy. In S. Hook (ed.), *Language and Philosophy*. New York: New York University Press. Reprinted in Chomsky (1972a).

Chomsky, N. (1970), Remarks on nominalization. In R. Jacobs and E. Rosenbaum (eds.), *Readings in English Transformational Grammar*. Waltham, Mass.: Ginn, 184–221.

Chomsky, N. (1971), *Problems of Knowledge and Freedom*. New York: Pantheon.

Chomsky, N. (1972a), *Language and Mind*. Enlarged edition. New York: Harcourt Brace Jovanovich.

Chomsky, N. (1972b), Some empirical issues in the theory of transformational grammar. In S. Peters (ed.), *Goals of Linguistic Theory*. Englewood Cliffs, N.J.: Prentice Hall, 63–130.

Chomsky, N. (1976), *Reflections on Language*. London: Temple Smith.

Chomsky, N. (1977), *Essays on Form and Interpretation*. Amsterdam: North Holland.

Chomsky, N. (1979), *Language and Responsibility*. Sussex: Harvester Press.

Chomsky, N. (1980a), *Rules and Representations*. Oxford: Blackwell.

Chomsky, N. (1980b), On cognitive structures and their development. In M. Piattelli-Palmarini (ed.), *Language and Learning: The Debate between Jean Piaget and Noam Chomsky*. London: Routledge and Kegan Paul, 35–52.

Chomsky, N. (1980c), Rules and representations. *Behavioral and Brain Sciences*, 3/1, 1–15.

Chomsky, N. (1981a), *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, N. (1981b), Principles and parameters in syntactic theory. In N. Hornstein and D. Lightfoot (eds.), *Explanations in Linguistics*. London: Longman.

Chomsky, N. (1982), *Some Concepts and Consequences of the Theory of Government and Binding*. Cambridge, Mass.: MIT Press.

Chomsky, N. (1986a), *Knowledge of Language: Its Nature, Origin and Use*. New York: Praeger.

Chomsky, N. (1986b), *Barriers*. Cambridge, Mass.: MIT Press.

Chomsky, N. (1987), Transformational Grammar: past, present, and future. Talk given at Kyoto, January.

Chomsky, N. (1988), *Language and Problems of Knowledge: The Managua Lectures*. Cambridge, Mass.: MIT Press.

Chomsky, N. (1990), Language and mind. In D. H. Mellor (ed.), *Ways of Communicating*. Cambridge: Cambridge University Press, 56–80.

Chomsky, N. (1991a), Some notes on economy of derivation and representation. In R. Freidin (ed.), *Principles and Parameters in Comparative Grammar*. Cambridge, Mass.: MIT Press, 417–545.

Chomsky, N. (1991b), Linguistics and adjacent fields: a personal view. In A. Kasher (ed.), *The Chomskyan Turn*. Oxford: Blackwell, 5–23.

Chomsky, N. (1991c), Linguistics and cognitive science: problems and mysteries. In A. Kasher (ed.), *The Chomskyan Turn*. Oxford: Blackwell, 26–53.

Chomsky, N. (1993), A Minimalist Program for linguistic theory. In K. Hale and S. J. Keyser (eds.), *The View from Building 20: Essays in Honor of Sylvain Bromberger*. Cambridge, Mass.: MIT Press, 1–52. Preliminary version in *MIT Occasional Papers in Linguistics* (1992), 1. Cambridge, Mass.: MITWPL.

Chomsky, N. (1995a), *The Minimalist Program*. Cambridge, Mass.: MIT Press.

Chomsky, N. (1995b), Bare phrase structure. In G. Webelhuth (ed.), *Government and Binding Theory and the Minimalist Programme*, Oxford: Blackwell, 383–440. Preliminary version in *MIT Occasional Papers in Linguistics* (1994), 5. Cambridge, Mass.: MITWPL.

Chomsky, N. (2000a), *The Architecture of Language*. New Delhi: Oxford University Press.

Chomsky, N. (2000b), *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.

Chomsky, N. (2000c), Minimalist inquiries: the framework. In R. Martin, D. Michaels and J. Uriagereka (eds.), *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*. Cambridge, Mass.: MIT Press, 89–155.

Chomsky, N. (2001a), Derivation by phase. In M. Kenstowicz (ed.), *Ken Hale: A Life in Language*. Cambridge, Mass.: MIT Press, 1–52.

Chomsky, N. (2001b), Beyond explanatory adequacy. *MIT Occasional Papers in Linguistics* 20. Cambridge, Mass.: MITWPL.

Chomsky, N. (2002), *On Nature and Language*. Cambridge: Cambridge University Press.

Chomsky, N. (2004a), *Hegemony or Survival: America's Quest for Global Dominance*. Harmondsworth: Penguin.

Chomsky, N. (2004b), Untitled talk given at *Tools in Linguistic Theory*, Budapest, May.

Chomsky, N. (2005a), Three factors in language design. *Linguistic Inquiry*, 36/1, 1–22.

Chomsky, N. (2005b), Language and the brain. In A. J. Saleemi, O-S. Bohn and A. Gjedde (eds.), *In Search of a Language for the Mind-Brain*. Aarhus: Aarhus University Press, 143–70.

Chomsky, N. (2005c), On phases. Unpublished ms., MIT.

Chomsky, N. and Halle, M. (1968), *The Sound Pattern of English*. New York: Harper and Row.

Chomsky, N. and Lasnik, H. (1977), Filters and Control. *Linguistic Inquiry*, 8, 425–504.

Chomsky, N. and Lasnik, H. (1993), Principles and parameters theory. In J. Jacobs, A. von Stechow, W. Sternefeld and T. Vennemann (eds.), *Syntax: An International Handbook of Contemporary Research*. Berlin: de Gruyter, 506–69.

Clahsen, H. and Muysken, P. (1986), The availability of universal grammar to adult and child learners – a study of the acquisition of German word order. *Second Language Research*, 2/2, 93–119.

Clark, H. H. and Clark, E. V. (1977), *Psychology and Language*. New York: Harcourt Brace Jovanovich.

Coleman, J. A. (1996), *Studying Languages: A Survey of British and European Students*. London: CILT.

Comrie, B. (1990), Second language acquisition and language universals research. *Studies in Second Language Acquisition*, 12/2, 209–18.

Cook, V. J. (1981), Some uses for second language learning research. *Annals of the New York Academy of Sciences*, 379, 251–8.

Cook, V. J. (1985), Chomsky's Universal Grammar and second language learning. *Applied Linguistics*, 6, 1–8.

Cook, V. J. (1990), Observational evidence and the UG theory of language acquisition. In I. Roca (ed.), *Logical Issues in Language Acquisition*. Dordrecht: Foris, 33–46.

Cook, V. J. (1991), The poverty-of-the-stimulus argument and multi-competence. *Second Language Research*, 7/2, 103–17.

Cook, V. J. (1993), *Linguistics and Second Language Acquisition*. Basingstoke: Macmillan.

Cook, V. J. (1994a), The metaphor of access to Universal Grammar. In N. Ellis (ed.), *Implicit Learning and Language*. London: Academic Press, 477–502.

Cook, V. J. (1994b), Universal Grammar and the learning and teaching of second languages. In T. Odlin (ed.), *Perspectives on Pedagogical Grammar*. Cambridge: Cambridge University Press, 25–48.

Cook, V. J. (2001), *Second Language Learning and Language Teaching*: Third edition. London: Edward Arnold.

Cook, V. J. (2002), Language teaching methodology and the L2 user perspective. In V. J. Cook (ed.), *Portraits of the L2 User*. Clevedon: Multilingual Matters, 325–44.

Cook, V. J. (2003), The poverty-of-the-stimulus argument and structure-dependency in L2 users of English. *IRAL*, 41, 201–21.

Cook, V. J. (2004), The spelling of the regular past tense in English: implications for lexical spelling and dual process models. In G. Bergh, J. Herriman and M. Mabärg (eds.), *An International Master of Syntax and Semantics: Papers Presented to Aimo Seppänen on the Occasion of his 75th Birthday*. Göteborg: Acta Universitatis Gothoburgensis, 59–68.

Cook, V. J., Long, J. and McDonough, S. (1979), The relationship of first and second language learning. In G. E. Perren (ed.), *The Mother Tongue and Other Languages in Education*. London: Centre for Information on Language Teaching and Research.

Cook, V. J., Iarossi, E., Stellakis, N. and Tokumaru, Y. (2003), Effects of the second language on the syntactic processing of the first language. In V. J. Cook (ed.), *Effects of the Second Language on the First*. Clevedon: Multilingual Matters, 214–33.

Cromer, R. F. (1987), Language growth with experience without feedback. *Journal of Psycholinguistic Research*, 16/3, 223–31.

Culicover, P. W. (1991), Innate knowledge and linguistic principles. *Behavioral and Brain Sciences*, 14/4, 615–16.

Denison, D. (1994), *English Historical Syntax*. Harlow: Longman.

Déprez, V. (1994), Underspecification, functional properties, and parameter setting. In B. Lust, M. Suner and J. Whitman (eds.), *Syntactic Theory and First Language Acquisition. Vol. 1*. Hillsdale, N.J.: Erlbaum.

Dulay, H. and Burt, M. (1973), Should we teach children syntax? *Language Learning*, 3, 245–57.

Epstein, S. D., Flynn, S. and Martahardjono, G. (1996), Second language acquisition: theoretical and experimental issues in contemporary research. *Brain and Behavioral Sciences*, 19/4, 746–52.

Eubank, L. (1996), Negation in early German-English interlanguage: more valueless features in the L2 initial state. *Second Language Research*, 12, 73–106.

Everett, D. (2005), Cultural constraints on grammar and cognition in Pirahã: another look at the design features of human language. *Current Anthropology*, 46/4, 621–46.

Fabbro, F. (1999), *The Neurolinguistics of Bilingualism: An Introduction*. Aylesbury: Psychology Press.

Felix, S. (1987), *Cognition and Language Growth*. Dordrecht: Foris.

Feyerabend, P. (1975), *Against Method*. London: Verso.

Fitch, W. T., Hauser, M. D. and Chomsky, N. (2005), The evolution of the language faculty: clarifications and implications. *Cognition*, 97/2, 179–210.

Flynn, S. and Lust, B. (2002), A minimalist approach to L2 solves a dilemma of UG. In V. J. Cook (ed.), *Portraits of the L2 User*. Clevedon: Multilingual Matters, 93–120.

Fodor, J. A. (1981), Some notes on what linguistics is about. In N. Block (ed.), *Readings in the Philosophy of Psychology*, London: Methuen, 197–207.

Fodor, J. A. (1983), *The Modularity of Mind*. Cambridge, Mass.: MIT Press.

Freidin, R. (1991), Linguistic theory and language acquisition; a note on structure-dependence. *Behavioral and Brain Sciences*, 14/4, 618–19.

Gardner, B. T. and Gardner, D. A. (1971), Two way communication with an infant chimpanzee. In A. Schrier and F. Stollnitz (eds.), *Behavior of Non-Human Primates: Vol. 4*. New York: Academic Press.

Gazdar, G. (1987), Generative grammar. In J. Lyons, R. Coates, M. Deuchar and G. Gazdar (eds.), *New Horizons in Linguistics 2*. London: Penguin, 122–51.

Gazdar, G. and Mellish, C. (1989), *Natural Language Processing in Prolog*. Wokingham: Addison Wesley.

Gazdar, G., Klein, E., Pullum, G. and Sag, I. (1985), *Generalised Phrase Structure Grammar*. Oxford: Blackwell.

Gentner, T. Q., Fenn, K. M., Margoliash, D. and Nusbaum, H. C. (2006), Recursive syntactic pattern learning by songbirds. *Nature*, 440, 1204–7.

Gleitman, L. (1982), Maturational determinants of language growth. *Cognition*, 10, 103–14.

Gleitman, L. (1984), Biological predispositions to learn language. In P. Marler and H. Terrace (eds.), *The Biology of Learning*. New York: Springer, 553–84.

Gold, E. M. (1967), Language identification in the limit. *Information and Control*, 10, 447–74.

Goodluck, H. (1986), Language acquisition and linguistic theory. In P. Fletcher and M. Garman (eds.), *Language Acquisition: Studies in First Language Development*. Cambridge: Cambridge University Press, 49–68.

Gopnik, M. and Crago, M. B. (1991), Familial aggregation of a developmental language disorder. *Cognition*, 39, 1–50.

Gordon, R. G., Jr. (ed.) (2005), *Ethnologue: Languages of the World*. Fifteenth edition. Dallas, Tex.: SIL International. Online version: www.ethnologue.com.

Gould, S. J. (1993), *Eight Little Piggies*. London: Jonathan Cape.

Guasti, M. T. (2002), *Language Acquisition: The Growth of the Grammar*. Cambridge, Mass.: Bradford Books.

Haegeman, L. (1990), Non-overt subjects in diary contexts. In J. Mascaró and M. Nespor (eds.), *Grammar in Progress*. Dordrecht: Foris, 167–74.

Haegeman, L. (1994), *Introduction to Government and Binding Theory*. Second edition. Oxford: Blackwell.

Haegeman, L. and Ihsane, T. (2002), Adult null subjects in the non-pro drop languages: two diary dialects. *Language Acquisition*, 9/4, 329–46.

Hale, K. and Keyser, S. J. (1993), On argument structure and the lexical expression of syntactic relations. In K. Hale and S. J. Keyser (eds.), *The View from Building 20: Essays in Honor of Sylvain Bromberger*. Cambridge, Mass.: MIT Press, 53–109.

Halliday, M. A. K., McIntosh, A. and Strevens, P. (1964), *The Linguistic Sciences and Language Teaching*. London: Longman.

Han, Z. (2002), A study of the impact of recasts on tense consistency in L2 output. *TESOL Quarterly*, 36, 543–72.

Hannan, M. (2004), A study of the development of the English verbal morphemes in the grammar of 4–9 year old Bengali-speaking children in the London borough of Tower Hamlets. PhD, University of Essex.

Hauser, M. D., Chomsky, N. and Fitch, W. T. (2002), The faculty of language: what is it, who has it, and how did it evolve? *Science*, 298, 1569–79.

Hawkins, J. A. (2003), *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.

Hawkins, R. and Chen, C. (1997), The partial availability of Universal Grammar in second language acquisition: the 'failed functional features' hypothesis. *Second Language Research*, 13/3, 187–226.

Haynes, J. H. (1971), *Morris Minor 1000 Owner's Workshop Manual*. Sparkford: Haynes Publications.

Hirsh-Pasek, K., Treiman, R. and Schneiderman, M. (1984), Brown and Hanlon revisited: mothers' sensitivity to ungrammatical forms. *Journal of Child Language*, 11/1, 81–8.

Hoban, R. (1967), *The Mouse and his Child*. New York: Harper and Row.

Hornstein, N., Nunes, J. and Grohmann, K. K. (2005), *Understanding Minimalism*. Cambridge: Cambridge University Press.

Horrocks, G. (1987), *Generative Grammar*. Harlow: Longman.

Howe, C. (1981), *Acquiring Language in a Conversational Context*. London: Academic Press.

Huang, C-T. J. (1984), On the distribution and reference of empty pronouns. *Linguistic Inquiry* 15, 531–74.

Hunston, S. and Francis, G. (2000), *Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins.

Hyams, N. (1986), *Language Acquisition and the Theory of Parameters*. Dordrecht: Reidel.

Hyams, N. (1992), A reanalysis of null subjects in child language. In J. Weissenborn, H. Goodluck and T. Roeper (eds.), *Theoretical Issues in Language Acquisition*. Hillsdale, N.J.: Erlbaum, 269–300.

Hyams, N. and Wexler, K. (1993), On the grammatical basis of null subjects in child language. *Linguistic Inquiry*, 24/3, 421–59.

Hymes, D. (1972), Competence and performance in linguistic theory. In R. Huxley and E. Ingram (eds.), *Language Acquisition: Models and Methods*. New York: Academic Press, 269–93.

Jackendoff, R. (1977), *X'-Syntax: A Study of Phrase Structure*. Cambridge, Mass.: MIT Press.

Jackendoff, R. (2002), *Foundations of Language*. Oxford: Oxford University Press.

Jaeggli, O. (1986), Passive. *Linguistic Inquiry*, 17, 587–622.

Jaeggli, O. and Safir, K. J. (1989), The null subject parameter and parametric theory. In O. Jaeggli and K. J. Safir (eds.), *The Null Subject Parameter*. Dordrecht: Kluwer, 1–44.

Kayne, R. (1984), *Connectedness and Binary Branching*. Dordrecht: Foris.

Kayne, R. (1994), *The Antisymmetry of Syntax*. Cambridge, Mass.: MIT Press.

Keenan, E. L. and Comrie, B. (1977), Noun phrase accessibility and universal grammar. *Linguistic Inquiry*, 8, 63–99.

Kegl, J. (1994), The Nicaraguan sign language project: an overview. *Signpost*, 7/1, 24–31.

Kenstowicz, M. (1994), *Phonology in Generative Grammar*. London: Edward Arnold.

King, E. S. (1980), *Speak Malay*. Kuala Lumpur: Eastern Universities Press.

Klein, W. and Perdue, C. (1997), The basic variety (or: couldn't natural languages be much simpler?). *Second Language Research* 13/4, 301–47.

Koopman, H. (1984), *The Syntax of Verbs*. Dordrecht: Foris.

Koopman, H. and Sportiche, D. (1991), The position of subjects. *Lingua*, 85, 211–58.

Lado, R. (1964), *Language Teaching: A Scientific Approach*. New York: McGraw-Hill.

Larsen-Freeman, D. and Long, M. (1991), *An Introduction to Second Language Acquisition Research*. London and New York: Longman.

Larson, R. (1988), On the double object construction. *Linguistic Inquiry*, 19, 335–91.

Legate, J. A. and Yang, C. D. (2002), Empirical reassessment of stimulus poverty arguments. *Linguistics Review*, 19, 151–62.

Lenneberg, E. H. (1967), *Biological Foundations of Language*. New York: Wiley.

Lightfoot, D. (1982), *The Language Lottery: Toward a Biology of Grammars*. Cambridge, Mass.: MIT Press.

Lightfoot, D. (1989), The child's trigger experience: degree-0 learnability. *Behavioral and Brain Sciences*, 12/2, 321–34.

Lightfoot, D. (1991), *How to Set Parameters: Arguments from Language Change*. Cambridge, Mass.: MIT Press.

Malinowski, B. (1923). The problem of meaning in primitive languages. In C. K. Ogden and I. A. Richards (eds.), *The Meaning of Meaning*. London: Routledge and Kegan Paul; 296–336.

McCawley, J. D. (1992), A note on auxiliary verbs and language acquisition. *Journal of Linguistics*, 28/2, 445–52.

McCloskey, J. (2000), Quantifier float and wh-movement in an Irish English. *Linguistic Inquiry* 31, 57–84.

McCloskey, J. (2002), Resumption, successive cyclicity, and the locality of operations. In S. Epstein and D. Seeley (eds.), *Prospects for Derivational Explanation*. Oxford: Blackwell, 184–226.

McDaniel, D. and Cairns, H. S. (1990), The child as informant: eliciting linguistic intuitions from young children. *Journal of Psycholinguistic Research*, 19/5, 331–44.

McNeill, D. (1966), Developmental psycholinguistics. In F. Smith and G. A. Miller (eds.), *The Genesis of Language: A Psycholinguistic Approach*. Cambridge, Mass.: MIT Press, 15–84.

Mehta, V. (1971), *John is Easy to Please*. London: Secker and Warburg.

Morgan, J. L. (1986), *From Simple Input to Complex Grammar*. Cambridge, Mass.: MIT Press.

Naoi, K. (1989), Structure-dependence in second language acquisition. *MITA Working Papers* (Keio University), 2, 65–77.

Nassaji, H. and Fotos, S. (2004), Current developments in research on the teaching of grammar. *Annual Review of Applied Linguistics*, 24, 126–45.

Nelson, K. E., Carskaddon, G. and Bonvillain, J. D. (1973), Syntactic acquisition: impact of experimental variation in adult verbal interaction with the child. *Child Development*, 44, 497–504.

Newmeyer, F. J. (2003), Grammar is grammar and usage is usage. *Language*, 9, 682–707.

Newport, E. L. (1977), Motherese: the speech of mothers to young children. In N. Castellan, D. Pisoni and G. Potts (eds.), *Cognitive Theory: Vol. 2*. Hillsdale, N.J.: Erlbaum, 177–217.

Ochs, E. and Schieffelin, B. (1984), Language acquisition and socialisation: three developmental stories and their implications. In R. Shweder and R. Levine (eds.), *Culture and its Acquisition*. New York: Cambridge University Press, 276–320.

Ouhalla, J. (1999), *Introducing Transformational Grammar: From Rules to Principles and Parameters*. London: Edward Arnold.

Paivio, A. and Begg, I. (1981), *Psychology of Language*. New York: Prentice Hall.

Pallier, C., Dehaene, S., Poline, J-B., Lbihan, D., Argenti, A-M. and Dupoux, E. (2003), Brain imaging of language plasticity in adopted adults: can a second language replace the first? *Cerebral Cortex*, 13, 155–61.

Paradis, M. (2004), *A Neurolinguistic Theory of Bilingualism*. Amsterdam: John Benjamins.

Patterson, F. G. (1981), Innovative uses of language by a gorilla: a case study. In K. Nelson (ed.), *Children's Language: Vol. 2*. New York: Gardner, 497–561.

Piaget, J. (1980), The psychogenesis of knowledge and its epistemological significance. In M. Piattelli-Palmarini (ed.), *Language and Learning: The Debate between Jean Piaget and Noam Chomsky*. London: Routledge and Kegan Paul, 23–34.

Piattelli-Palmarini, M. (ed.) (1980), *Language and Learning: The Debate between Jean Piaget and Noam Chomsky*. London: Routledge and Kegan Paul.

Pinker, S. and Jackendoff, R. (2005), The faculty of language: what's special about it? *Cognition*, 95/2, 201–36.

Poeppel, D. and Wexler, K. (1993), The full competence hypothesis of clause structure in early German. *Language*, 69, 1–33.

Pollard, C. and Sag, I. A. (1994), *Head-Driven Phrase Structure Grammar*. Stanford: Stanford University Press.

Pollock, J-Y. (1989), Verb movement, universal grammar, and the structure of IP. *Linguistic Inquiry*, 20, 365–424.

Pullum, G. K. and Scholz, C. (2002), Empirical assessment of stimulus poverty arguments. *Linguistic Review*, 19/1, 9–50.

Raposo, E. (1987), Case Theory and Infl-to-Comp: the inflected infinitive in European Portuguese. *Linguistic Inquiry* 18, 85–109.

Rizzi, L. (1982), *Issues in Italian Syntax*. Dordrecht: Foris.

Rizzi, L. (1986), Null objects in Italian and the theory of *pro*. *Linguistic Inquiry*, 17, 501–57.

Rizzi, L. (1990), *Relativised Minimality*. Cambridge, Mass.: MIT Press.

Rizzi, L. (1997), On the fine structure of the left-periphery. In L. Haegeman (ed.), *Elements of Grammar*, Dordrecht: Kluwer, 281–337.

Rizzi, L. (2004), On the study of the language faculty: results, developments and perspectives. *Linguistic Review*, 21, 323–44.

Roca, I. (1994), *Generative Phonology*. London: Routledge.

Roeper, T. (1999), Universal bilingualism. *Bilingualism: Language and Cognition*, 2, 169–86.

Ross, J. R. (1967), Constraints on variables in syntax. PhD, MIT.

Rumbaugh, D. M. (ed.) (1977), *Language Learning by a Chimpanzee: The Lana Project*. New York: Academic Press.

Rumelhart, D. E. and McLelland, J. L. (1986), On learning the past tenses of English verbs. In J. L. McLelland, D. E. Rumelhart and the PDP Research Group, *Parallel Distributed Processing. Vol. 2: Psychological and Biological Models*. Cambridge, Mass.: MIT Press, 216–71.

Schachter, J. (1988), Second language acquisition and its relationship to Universal Grammar. *Applied Linguistics*, 9/3, 219–35.

Schwartz, B. and Sprouse, R. (1996), L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research*, 12/1, 40–72.

Senghas, A. and Coppola, M. (2001), Children creating language: how Nicaraguan sign language acquired a spatial grammar. *Psychological Science*, 12/4, 323–8.

Senghas, A., Kita, S. and Asli Özyürek, A. (2004), Children creating core properties of language: evidence from an emerging sign language in Nicaragua. *Science*, 305/5691, 1779–82.

Sharpe, T. (1982), *Vintage Stuff*. London: Secker and Warburg.

Siegel, D. (1974), Topics in English morphology. PhD, MIT.

Simmons, D. (2004), *Illium*. London: Gollancz.

Sinclair, J. and Coulthard, M. (1975), *Towards an Analysis of Discourse*. Oxford: Oxford University Press.

Skinner, B. F. (1957), *Verbal Behavior*. New York: Appleton Century Crofts.

Spolsky, B. (1989), *Conditions for Second Language Learning*. Oxford: Oxford University Press.

Stowell, T. (1981), Origins of phrase structure. PhD, MIT.

Surányi, B. (2004), Head movement qua Root Merger. In L. Varga (ed.), *The Even Yearbook 6. ELTE SEAS Working Papers in Linguistics*, Department of English Linguistics, Eötvös Loránd University, Budapest, 167–83.

Tomasello, M. (1999), *The Cultural Origins of Human Cognition*. Cambridge, Mass.: Harvard University Press.

Tomasello, M. (2003), *Constructing a Language*. Cambridge, Mass.: Harvard University Press.

Toro, J. M., Trobalon, J. B. and Sebastien-Galles, N. (2005), Effects of backward speech and speaker variability in language discrimination by rats. *Journal of Experimental Psychology: Animal Behaviour Processes*, 31/1, 95–100.

Travis, L. (1984), Parameters and effects of word order variation. PhD, MIT.

Vainikka, A. and Young-Scholten, M. (1996), Gradual development of L2 phrase structure. *Second Language Research*, 12/1, 7–39.

Wallman, J. (1992), *Aping Language*. Cambridge: Cambridge University Press.

Watt, W. C. (1967), Structural properties of the Nevada cattle-brands. *Computer Science Research Review* (Carnegie-Mellon University), 2, 20–7.

Wells, C. G. (1985), *Language Development in the Pre-School Years*. Cambridge: Cambridge University Press.

Wexler, K. and Culicover, P. W. (1980), *Formal Principles of Language Acquisition*. Cambridge, Mass.: MIT Press.

Wexler, P. and Manzini, R. (1987), Parameters and learnability in Binding Theory. In T. Roeper and E. Williams (eds.), *Parameter Setting*. Dordrecht: Reidel, 41–76.

White, L. (1991), Adverb placement in second language acquisition: some effects of positive and negative evidence in the classroom. *Second Language Research*, 1, 133–61.

White, L. (2003), *Second Language Acquisition and Universal Grammar*. Cambridge: Cambridge University Press.

Willis, J. (1996), *A Framework for Task-Based Learning*. Harlow: Longman.

Wode, H. (1981), *Learning a Second Language*. Tübingen: Narr.

Zuidema, W. (2003), How the poverty of the stimulus solves the poverty of the stimulus. In S. Becker, S. Thrun and K. Obermayer (eds.), *Advances in Neural Information Processing Systems 15*. Proceedings of NIPS '02. Cambridge, Mass.: MIT Press, 51–8.

Index

A-bar movement, *see* Ā-movement
A-bar position, *see* Ā-position
Abney, S. P., 105–6
abstract Case, 147
access to UG in second language acquisition, 231
Accessibility Hierarchy, 22–3
Accusative Case, 74–5
acquisition of first language, 12, 26, 45–60, 93, 185–220
active categories, 295
Adger, D., 1
Adjacency, 155–8
Adjective Phrase, 67
Adjunct, 78–80, 113, 125
adjunction, 137–8, 265–8
affix hopping, 132
agent θ-role, 81
AGR, *see* agreement
agreement (AGR), 91–2, 95–9, 101, 109, 111–12, 127, 130, 148–9, 153, 155, 157
Agreement Phrase, *see* AgrP
AgrP (Agreement Phrase), 109, 111–12
aims of linguistics, 10–12
American Sign Language (ASL), 187
A-movement 121–5, 127, 133–4, 145, 248
Ā-movement, 126–7, 133, 171
anaphors, 38, 114, 163, 166, 226, 248
apes and language, 47, 186–7
A-position, 125, 127, 135, 145, 171–2, 243–4
Ā-position, 127, 171, 244
Arabic, 40–1, 45, 91, 97, 214
arbitrary reference, 89
arguments, 81–5, 87, 122, 124–5, 127, 135, 145–6, 155–6, 160–1, 178
Articulated INFL Hypothesis, 112, 119
ASL, *see* American Sign Language
Aspects Model, 3, 15, 19
Atkinson, M., 54

Bahasa Malaysia, 21
Baker, C. L., 189
Baker, M., 122, 263, 265
Bald, W-D., 226
Bare Output Conditions, 248, 252, 254, 257, 262, 266
bare phrase structure, 257–9, 261–6, 270–1, 277
Barriers syntax, 139, 141–4, 302
Begg, I., 51
behaviourism, 51–2, 200
Bellugi, U., 200, 227
Berwick, R., 198, 215
Biber, D., 18
Bickerton, D., 188
bilinguals, 221–4, 230
Binary Branching Condition, 113, 255–6
binding principles, 166, 168, 196, 226
Binding Theory, 3, 72–3, 162–74, 182, 194, 267
biology and language, 185, 248
Birdsong, D., 238
Bley-Vroman, R., 233
Blocking Category, 140, 143–4
Bloomfield, L., 13, 51
Borer, H., 98, 208, 283
Botha, R., 35
bounding nodes, 139, 140–2
Bounding Theory, 72–3, 76, 84, 90, 138–41, 143–4, 162, 171, 181

bracketed input, 193
brackets, 29, 31
brain and language, 12–13, 25
Braine, M., 192, 207
broad faculty of language (FLB), 48, 216
Brown, R., 198–202, 227
Bruner, J., 193, 201–3
Burzio, L., 160–1
Burzio's generalization, 161

C, *see* complementizer
C″, *see* Complementizer Phrase
Cairns, H. S., 216
Carroll, S., 225
Case, 74–6, 146–53, 161–2, 168, 174, 182–4, 248, 263, 276–8, 282, 284–5, 289–90, 292–6, 299, 309
Case assigner, 76, 147–50, 153–60
Case Filter, 75–6, 147, 191, 195, 197, 203–4, 228, 278, 294
Case Theory, 76, 146–61
causative verbs, 132, 137, 264
Cazden, C. R., 201
c-command (constituent command), 114–15, 165–6, 269
chains, 135, 138, 247
Checking Theory, 276–8, 283–5, 290–5, 306–7
Chinese, 45, 95–6, 180, 213
Chomsky, N.: for particular models, *see Aspects* Model; *Barriers* syntax; Government/Binding Theory; Minimalist Program; Principles and Parameters Theory; Standard Theory; *Syntactic Structures* model; transformational generative grammar; for particular concepts, *see* biology and language; brain and language; broad faculty of language; competence; computational system; conceptual-intentional system; core grammar; environment and language; explanatory adequacy; externalized language; generative grammar; 'growth' of language; innate aspects of language; Language Acquisition Device; maturation; narrow faculty of language; perfection of language; performance; Plato's problem; positive evidence in language acquisition; poverty-of-the-stimulus argument; pragmatic competence; psychology and linguistics; uniformity requirement
Chomsky and politics, 1, 27
Clahsen, H., 232
cognate object, 123
cognitive development and language, 202–3, 220, 227
Coleman, J. A., 225
communication, 16–17
communicative competence, 15–16
competence, 4, 15–16, 19, 186, 192–3, 196–7, 204–5, 207, 215–16, 218–19, 221–4, 229–31, 233, 238
complement, 41–3, 44, 65–6, 69, 74–5, 78, 83, 85, 102, 105, 113, 123, 151, 153, 155–6, 252, 261, 263–6, 268, 303
complementizer (C), 93–4, 100–4, 126–7, 130, 149–51, 153, 157, 159, 175–7, 273, 285, 298, 305, 307
Complementizer Phrase (C″/CP), 100, 102–4, 109, 112, 118, 126–8, 134, 136, 140–1, 144, 271, 290, 301, 302, 307
computational procedure, 149–51, 153, 156–7, 175–7, 273, 275, 277, 285, 298–301, 303, 305, 307
computational system, 3, 5, 6, 8–11, 248–9
conceptual-intentional system, 5, 9–10, 12, 48, 248
connectionism, 25
constituent command, *see* c-command
Control Theory, 86–90, 119, 173
controller, 89–90
convergence, 252–4
Cook, V., 25, 56–7, 209, 217, 221, 224–5, 230–5, 237, 239–40
Coppola, M., 188
Copy-Delete movement, 279–81
core grammar, 25–6, 59, 185, 187, 190, 203, 206, 214–15, 219, 227
correction of learner's speech, 196–9, 226–7
Coulthard, M., 227
covert movement, 281–4, 307
CP, *see* Complementizer Phrase
CPH, *see* Critical Period Hypothesis
Crago, M. B., 186
creativity, 17–18, 51, 187–8, 195–6, 202–3
Creoles, 188
Critical Period Hypothesis (CPH), 237
Cromer, R., 206
Culicover, P. W., 194

D, *see* Determiner
deep structure, 3–4, 61, 64, 119; *see also* D-structure
degeneracy of the data, 192–3
degree-0 learning, 194
Denison, D., 194
descriptive adequacy, 54–5
Det, *see* Determiner
Determiner (D/Det), 29, 107
Determiner Phrase (DP), 105–9, 122, 125, 127–8, 133, 141–2, 147–9, 151–2, 155–6, 158–9, 161–2, 164–6, 168–71, 174, 180, 183–4
development of language, 46, 49–50, 185, 192, 200, 202, 211, 215–19, 223, 230, 236–7
direct access to UG in second language acquisition, 74, 231
do-insertion, 129, 247
dominate, 83, 100, 114, 144, 268
DP, *see* Determiner Phrase
DP Hypothesis, 105–8, 256–7
D-structure, 3–4, 61, 121, 124–5, 134–5, 148, 159, 161, 183, 249, 253, 262–3; *see also* deep structure
Dutch, 91, 99

ECM, *see* Exceptional Case-Marking
economy of derivation, 247, 249, 251
economy of grammar, 248–9
economy of representation, 246, 249
ECP, *see* Empty Category Principle
E-language, *see* externalized language
elegance, 248
empty (e) categories, 134–5, 170
Empty Category Principle (ECP), 92, 175, 177–8, 181, 183–4, 247, 275–6, 287
English: for particular aspects of English, *see* bounding nodes; Exceptional Case-Marking; head parameter; non-pro-drop languages; PRO; Subjacency
environment and language, 13, 50, 56, 58, 185–8, 191–3, 203, 205–6, 222, 224, 234
EPP, *see* Extended Projection Principle
EPP feature, 293–4, 296–301, 303, 305–8
Epstein, S., 235
ergative verbs, 116
Eubank, L., 238
Everett, D., 48
evidence for language acquisition, 189–94, 215, 224–6, 228, 232, 234–6
Exceptional Case-Marking (ECM), 153–5, 292
expansion of children's utterances, 200–1, 217
experiencer θ-role, 82
explanations to learners, 196, 226
explanatory adequacy, 54–5, 121
expletive (pleonastic) subjects, 86, 92, 125, 212
Extended Projection Principle (EPP), 87, 90, 134, 169, 213, 293–4, 296–8, 299–301, 303, 305–8; *see also* EPP feature
external arguments, 83, 85
External Merge, 271–5, 281–2, 287
externalized (E-)language, 4, 6, 13–15, 17–18

Fabbro, F., 12
Failed Functional Features Hypothesis, 237
Farsi, *see* Persian
feature movement, 284
Feyerabend, P., 60
FI Principle, *see* Full Interpretation Principle
final L2 state (S_f), 230–1, 238–41
Finnish, 99
Fitch, W. T., 48
FLB, *see* broad faculty of language
FLN, *see* narrow faculty of language
Flynn, S., 235
Fodor, J. A., 36, 46
Fotos, S., 240
French, 39–40, 91, 110–11, 129–31, 239
Full Access Hypothesis, 232–3, 235
Full Access/Full Transfer Hypothesis, 235
Full Competence Hypothesis, 208, 217
Full Interpretation (FI) Principle, 245–6, 249, 252
Functional Parameterization Hypothesis, 283
functional phrases, 100–12

Gall, F., 48
Gardner, B. T., 187
Gardner, D. A., 187
Gazdar, G., 35
GB Theory, *see* Government/Binding Theory
generative grammar, 2, 13, 35–6, 52, 164
Genitive case, 75, 147, 152

Copyrighted Material

Gentner, T., 186
German, 39, 47, 91, 94, 99, 129–30, 232, 236
Gleitman, L., 192, 218
goal θ-role, 81, 85
Gold, E. M., 189, 194
Goodluck, H., 218
Gopnik, M., 186
governing category, 167–8, 173
Government/Binding (GB) Theory, 1, 3–4, 35, 62, 69, 73, 75, 93, 121, 185, 242–3; *see also* Binding Theory; Bounding Theory; Case Theory; Government Theory; proper government; Theta Theory; X-bar (X′) theory
Government Theory, 88, 91, 140–1, 152
governors, 88, 150–1
gradual development hypothesis, 208, 217, 236
grammar, 7, 49
Greek, 91, 239
'growth' of language, 49, 185, 217–18
Guasti, M. T., 185

Haegeman, L., 54, 96, 222
Hale, K., 263
Han, Z., 227
Hanlon, C., 199–200
Hannan, M., 216
Hauser, M. D., 48, 119
Hawkins, J., 44
Head-Driven Phrase Structure Grammar (HPSG), 35
head-first (initial)/head-last (final), *see* head parameter
head movement, 128, 130, 133, 288–90
Head Movement Constraint, 145–6, 204, 226
head of phrase, 41, 69, 74, 258
head parameter, 41–5, 50, 54, 59, 65, 74, 198, 205, 213, 269
Hebrew, 98
Hirsh-Pasek, K., 199–201
Hornstein, N., 1
Howe, C., 199
HPSG, *see* Head-Driven Phrase Structure Grammar
Huang, J-T., 96, 99
human species, 47
Hungarian, 33–4, 40, 44, 88, 90, 92–4, 98, 105, 109–11, 132–3, 172, 179
Hyams, N., 211–13, 216–17
Hymes, D., 16

I, *see* inflection
Icelandic, 99
I-language, *see* internalized language
imitation as a method of language acquisition, 195–6, 201, 203, 206, 218, 225, 227–8
inactive categories, 295
indices, 164–5
indirect access to UG in second language acquisition, 231
INFL, *see* inflection
inflection (INFL/I), 101, 112, 119, 127
Inflection Phase (IP), 101–4, 243–4, 271–3, 283, 290, 292–6, 298, 301, 303, 305–6
inherent Case, 148, 154–5, 158, 161, 184
initial L2 state (S_0), 205, 230–7
innate aspects of language, 2, 26, 49, 56–9, 204–6, 210, 218
input to language acquisition, 53–4, 56–8, 188–9, 192–4, 206, 217, 220, 225, 228–9, 231, 240
instrument (θ-)roles, 82–3
interface conditions, 4, 248
interface levels, 4, 6–8, 10, 12, 48, 248–9, 252, 254–5, 262
internal argument, 83, 85
Internal Merge, 271–2, 274
internalized (I-)language, 6, 13–17, 26, 49
IP, *see* Inflection Phase
Irish, 300
islands, 71, 73, 139–43, 298, 301–2
Italian, 45, 92–3, 95, 123, 210, 212–13

Jackendoff, R., 7, 48, 69, 256
Jaeggli, O., 95, 99, 160
Japanese, 5, 21, 34, 42–4, 91, 95, 186, 190–1, 210, 215, 234–5, 239

Kayne, R., 239, 255, 269
Kenstowicz, M., 7
Keyser, S. J., 263
Klingon, 120
Koopman, H., 84, 133, 156
Korean, 186

L2 acquisition, *see* second language acquisition

LAD, *see* Language Acquisition Device
Lado, R., 225
Language Acquisition Device (LAD), 52–5, 206, 228–9; *see also* language faculty; Universal Grammar Theory
Language Acquisition Support System (LASS), 193
language faculty, 14, 23, 45–50, 54, 57, 60; *see also* broad faculty of language; narrow faculty of language
Larson, R., 114, 116, 118, 266
Lasnik, H., 175–6, 276
LASS, *see* Language Acquisition Support System
Latin, 74
learning strategies, 195
legibility conditions, 12, 252
Lenneberg, E., 185–6
lexical entries, 8, 54, 61, 65, 82, 124
Lexical Parameterization Hypothesis, 283
lexicon, 8–10, 13, 29, 32, 59, 61, 76–7
LF, *see* Logical Form
light verbs, 117, 264–5
Lightfoot, D., 46, 194
linear order, 268–70
L-marking, 143–4
local domain, 166–8
Locality Principle, 21, 23–4, 36–41, 45, 48–9, 55–7, 70, 281, 290, 295–300
Logical Form (LF), 7–8, 10, 180, 183, 247, 252, 268
Lust, B., 235

Malagasy, 31
Malinowski, B., 17
Manzini, R., 215
markedness, 215–20
maturation, 185, 207, 217–20
maximal projections, 67–9, 78, 101–2
McCawley, J. D., 53
McNeill, D., 200
m-command, 150–2
mental organ, 46–7, 49, 186
Merge, 78, 249, 252, 254–5, 262, 265, 267, 271–2
middle verbs, 122–3, 125, 161, 239
Minimal Link Condition, 247, 295
Minimal Trees Hypothesis, 236
Minimalist Program (MP), 1, 3, 8, 78, 118–19, 219, 242–306
modules of GB, 46, 62–72, 121, 138–9, 159
monolingual native speaker, 221–2, 224, 229, 238
Morgan, J., 193–4
morphological uniformity, 96–8
mother, 83
Move α, 137–8, 247
Move-Insert, 279
movement, 20–1, 32–3, 36, 61–2, 69–73, 76, 90, 94–5, 121–84, 194, 205, 215, 230, 234–5, 237, 239, 244, 246–7, 269, 271–309
MP, *see* Minimalist Program
multi-competence, 223–4, 239
Muysken, P., 232

N″, *see* Noun Phrase
Naoi, K., 234
narrow faculty of language (FLN), 48, 216
Nassaji, H., 240
negative evidence, 190–1, 195, 197–8, 201–4, 210, 212–13, 220, 227
negative fronting, 127
negative polarity, 115
Nelson, K. E., 201
Newmeyer, F., 19
Newport, E. L., 192
Nicaraguan Sign Language, 188
No Merger, 268
No Tamper Condition, 287–91
No UG Hypothesis, 232
Nominative Case, 74–5
non-finite clauses, 87
non-pro-drop languages, 91–4, 99, 120, 191, 205, 210–15
Norwegian, 94–5
Noun Phrase (N″/NP), 28
NP, *see* Noun Phrase
NP movement, 162; *see also* A-movement; movement
null objects, 96
null subject, *see* pro-drop parameter
Numeration, 251–5, 257–8, 260, 270–1, 276

observational adequacy, 54–5
occurrence requirement, 190, 196, 225–8
Old English, 96, 194
organ of language, 46–7, 49
overt movement, 281–4

P, *see* Prepositions
P″/PP, *see* Preposition Phrase

Paivio, A., 51
Paradis, M., 12
parameter setting, 59, 93
parameters, 3, 9, 21, 44; *see also* head parameter; pro-drop parameter
Partial Access Hypothesis, 236
passive constructions, 2, 9, 20, 33, 80, 117, 122, 125, 161, 296
past tense '*-ed*', 25, 59, 98–9, 109, 137
patient θ-role, 81–3, 85, 263
Pattern Grammar, 13
Patterson, F. G., 187
perfection of language, 3, 12, 248
performance, 15, 18–20, 26
peripheral grammar, 25–6
Persian, 99
PF, *see* Phonetic Form
Phases Model, 4, 301–8
Phonetic Form (PF), 7, 10, 178, 252, 267, 269
phrase structure, 28, 31, 35, 61–6, 69, 255–61; *see also* X-bar theory
Piaget, J., 46–7, 203
Pinker, S., 48
Plato's problem, 55–6, 204
pleonastic subjects, *see* expletive subjects
Pollock, J-Y., 110, 112, 239, 246, 282
positive evidence in language acquisition, 187, 190–1, 195–7, 202–4, 210, 212–13, 215, 225, 228
postpositions, 42
poverty-of-the-stimulus argument, 55–7, 189, 224–8; *see also* Plato's problem
P & P Theory, *see* Principles and Parameters Theory
pragmatic competence, 16, 19, 202–3, 227, 240
precedence, 268
predicates, 81–2, 115, 119
Preposition (P), 42, 75
Preposition Phrase (P″/PP), 42, 65, 67, 69, 127, 156
primary linguistic data, 52, 54, 57
principles, 3, 8, 26, 36, 41, 45; *see also* binding principles; Empty Category Principle; Extended Projection Principle; Full Interpretation Principle; Locality Principle; Projection Principle; Structural Preservation Principle; Subjacency; Subset Principle; X-bar principles
Principles and Parameters Theory (P & P Theory), 3, 10, 24, 28, 34–5, 39, 41, 44, 54, 58, 62, 242
PRO (big PRO), 88–91, 119, 124, 134, 147, 150, 152, 172–4
pro (little pro), 91, 94, 98, 172
PRO Theorem, 88–9, 172–3
probes, 295–6, 298, 300, 303
Procrastinate, 282–4
pro-drop languages, 91, 94–5, 123, 172, 210–11, 214
pro-drop parameter, 86, 96, 99–100, 205, 209–15, 240
Projection Principle, 65–6, 77–8, 137–8
pronominal anaphors, *see* PRO
pronominals, 163, 167; *see also* pronouns
Pronouns, 38, 75, 87, 89–91, 94, 96, 124, 152, 162–4, 167–8, 172, 257, 267
proper government, 91–2, 99, 177
psychology and linguistics, 9, 24, 46, 52, 186
Pullum, G., 58

raising, 71, 159, 161
raising verbs, 124–6, 160
Raposo, E., 149
recipient θ-role, 81–2, 86
reciprocals, *see* anaphors
recursion, 18, 48–9, 79
referring expression, *see* r-expression
reflexives, *see* anaphors
relative clauses, 9, 22–3, 40, 128, 267
Relativized Minimality, 139, 145–6, 151–2, 247
remnant movement, 289
rewrite rules, 2, 4, 28–32, 35, 63, 66, 245
r-expression (referring expression), 163–4, 167, 267
Rizzi, L., 57, 59, 91–9, 99, 118, 141, 145–6, 151
Roca, I., 7
Roeper, T., 222–3
Ross, J. R., 71, 139–40, 142, 301
rules, 2–3, 4, 9–10, 14, 18, 24, 28
Rumbaugh, D. M., 187

S, *see* sentence
Safir, K., 95, 99
Scholz, C., 58
Schwartz, B., 235
scientific theory, 23, 27, 36

Copyrighted Material

second language (L2) acquisition, 25, 55, 197, 221–41
Senghas, A., 188
sensorimotor system, 5, 9, 12, 48
sentence (S), 1–2, 4, 13–15, 17–18, 23, 25, 29, 31, 101
Sinclair, J., 227
single bar (X′) category, *see* X-bar theory
Sisterhood Condition, 84–5, 113, 124
sisters, 79, 83, 113–14, 256–8, 260, 263, 265
Skinner, B. F., 51
social interaction, 14, 17, 201–2, 227
Spanish, 91, 210–12, 224, 235, 236
spec, *see* specifier
species and language, 47–9
specifier (spec), 66, 68–9, 105, 125, 244
Spell Out, 252–5, 284, 291–2, 302, 304
Split-INFL hypothesis, 109–13
Spolsky, B., 238
Sportiche, D., 84
Sprouse, R., 235
S-structure, 3–4, 61, 121, 135, 180; *see also* surface structure
Standard Theory, 3–4, 23, 62, 103
states of the language faculty, 49, 230; *see also* final L2 state; steady state
steady state (S_s), 50–1, 205, 230, 232
Stowell, T., 151
structural Case, 148, 154–5
Structural Preservation Principle, 134
structure-dependency, 234–5
subcategorization, 65
Subjacency, 139–43, 301
subject raising, 37, 161
subject–auxiliary inversion, 37, 128, 133
Subset Principle, 215
substitution, 137
superiority effect, 177–8, 182
Surányi, B., 289
surface structure, 3–4, 61, 74–6, 93, 119, 122, 132, 180, 232; *see also* S-structure
Swain, M., 225
Syntactic Structures model, 2–3, 62–4

take-up requirement, 192
Tense Phrase (TP), 109, 111
TG/TGG, *see* transformational generative grammar
that-trace, 175–6
thematic (θ-)roles, 81, 86, 262–4
theme (θ-)role, 82
Theta (θ-)Criterion, 85–6, 123, 135, 239, 244
theta (θ-)grid, 82
theta (θ-)marking, 84, 148
theta (θ-)position, 125
Theta (θ-)Theory, 80–5, 119, 262, 265
Tomasello, M., 46, 60, 188, 202
topic-drop, 96–7
Toro, J. M., 186
TP, *see* Tense Phrase
trace (t), 134, 138, 172
Trace Theory, 133
traditional grammar, 30
transformational generative grammar (TG/TGG), 2, 4, 35, 62
transformations, 33–5, 70, 121–2, 132–4, 138, 140
Travis, L. M., 145
trees, 29, 31, 267–8
Tulloch, J., 26

UG Theory, *see* Universal Grammar Theory
unaccusative verbs, 122–3, 125, 161, 264
Uniform Theta-Role Assignment Hypothesis (UTAH), 122, 263–6
uniformity requirement, 191–3, 197, 225, 228
uninterpretable features, 276–9, 282, 284, 290, 293, 306–8
universal bilingualism, 222
Universal Grammar (UG) Theory, 1, 9, 21, 24–5, 49; *see also* acquisition of first language; core grammar; input to language acquisition; Language Acquisition Device; language faculty; organ of language; poverty-of-the-stimulus argument; states of the language faculty
Universal Word Order approach, 269
universals, 20, 22–3, 39–41, 45, 59, 110, 147, 158, 215, 224, 242, 263; *see also* Universal Grammar Theory
use of language 12–13, 16
UTAH, *see* Universal Theta-Role Assignment Hypothesis

V2, *see* Verb-Second
V″/VP, *see* Verb Phrase
Vainikka, A., 236
Valueless Features Hypothesis, 237

Copyrighted Material

Vata, 133
Verb Phrase (V″/VP), 28, 42, 64–5, 80
Verb-Second (V2), 129–30
Verb–Subject order, 193
VP-Internal Subject Hypothesis, 84–5, 104, 125
VP-Shell Hypothesis, 113, 116, 118–19, 125, 264, 266, 268

Wallman, J., 187
Weinberg, A., 198
Wells, C. G., 201
Welsh, 22
Wexler, K., 194, 213, 215, 217
wh-elements, 20, 24, 38, 71, 73
wh-Island Constraint, 71
White, L., 221, 226, 233, 238
wh-movement, 24, 38, 70, 277
wh-questions, 20, 177–8
Willis, J., 225

X′-movement, 257
X′ Theory, *see* X-bar Theory
X-bar principles, 76
X-bar (X′) Theory, 43, 65–7, 69–70, 73–9, 100–10, 119, 135, 159, 162, 181, 255–7

Young-Scholten, M., 236